VOLUME X NOVEMBER 1919 NUMBER 9

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

DERIVATION OF A SCALE TO MEASURE ABILITIES IN SCIENTIFIC THINKING*

JOHN P. HERRING

THE measurement of ability in scientific thinking within defined limits has been undertaken by means of a series of problems which appeared in THIS JOURNAL in December 1918. Statements were there made about purpose, structure, limitations, and value, and the history of the test was included. The present article concerns the technique employed in the development of scales for this and similar tests.

I. A PRINCIPLE AND TWO ASSUMPTIONS WHICH UNDERLIE THIS TYPE OF INVESTIGATION

Underlying the interpretation of mental tests is the principle that mental abilities are distributed as are physical traits. If 1000 men, selected at random, pass by in the order of their height, we shall see

first, a very few very short men,
then a greater number of short men,
then a very great number of men of average height,
then a lesser number of tall men,
then a very few very tall men.

If the same men pass by in the order of their ability to remember words, we shall see

first, a very few with very poor memory,
then a greater number with poor memory,
then a very great number with average memory,
then a lesser number with good memory,
then a very few with very good memory.

*This article is a continuation of the author's study *Measurements of Some Abilities in Scientific Thinking*, which is to be found in THIS JOURNAL for December, 1918, Vol. IX, pp. 535-558. The problems there presented should be kept in mind throughout the present discussion.

If the selection is really random, the distribution will take this normal form. "We see the measures distributed somewhat symmetrically about a single, central measure and decreasing in frequency as we pass from the central measure toward either extreme, slowly at first, then more rapidly, then more slowly." (Thorndike, 1913, page 32). Only in the case of an infinite number of measures randomly selected would this curve correspond perfectly to its formula of derivation. Finite groups show irregularities which tend to decrease as the number of measures increases.

There are two assumptions which underlie mental tests of this sort. First, the distribution of the abilities to be studied is normal; and second, the variability of each large randomly selected group of pupils is approximately the same as that of any other.

II. The Probability Curve and its P. E. as a Unit of Measure*

In order to prove the validity of such a test as this, the properties of the frequency curve must be understood. Between the curve and its base are, e. g., all the pupils who took a certain test in arithmetic, arranged according to the number of problems each solved. The height of the curve at a given point (ordinate) represents the number of pupils who solved the number of problems directly below upon the base. Pupils are represented by equal amounts and per cents of area.

The most usual expression of the measures of individuals and of groups has been the per cent, but decimal expression is not always the best. Per cents state exactly the proportion of each group solving each problem and they indicate exactly the order of difficulty, but they are inadequate to establish the exact amount of difficulty. This is for two reasons.

First, the difference between 50% and 60% is equal to that between 90% and 100% as an expression of the number and of the fraction of pupils lying between, but not as an expression of increments of difficulty. In terms of difficulty, per cents lying near 50% are small, while those lying near 0% and 100% are large. It is very hard, not having any ability to spell, to acquire the first notions and the first spellings, and to pass from just not any spelling ability to

*Note: "The P. E. stands for Probable Error, the traditional, but very misleading, name for the median deviation, which is not especially probable, and not an error at all."—Thorndike, 1913, page 40, note.

10%; while it is easy to pass from 40% to 50% ability; but it is very hard again to pass from 90% to 100% or perfection,—that condition in which no words are ever missed and none need ever be studied. This means that as we pass from the left end of the probability curve to the median we see

first, 9.9% of area (or pupils) corresponding to 2.700 P. E.,
then 10% of area corresponding to .652 P. E.,
then 10% of area corresponding to .470 P. E.,
then 10% of area corresponding to .402 P. E.,
then 10% of area corresponding to .376 P. E.,

and as we pass from the median to the right end we see

first, 10% of area (or pupils) corresponding to .376 P. E.,
then 10% of area corresponding to .402 P. E.,
then 10% of area corresponding to .470 P. E.,
then 10% of area corresponding to .652 P. E.,
then 9.9% of area corresponding to 2.700 P. E.

The percent has often been used to measure mental abilities just as if it were a constant unit. It is a variable, and as such, it can be used safely only when the amounts and effects of its variations have been determined.

Second, for want of a common unit, per cents are not comparable from group to group of pupils, except roughly and with an unknown amount of error.

There is, however, a suitable unit for use instead of per cents,—the P. E., or Probable Error. It represents equal steps of difficulty within a group. It is comparable from group to group whenever it can be shown or properly assumed that the groups are approximately, and that their variabilities are approximately equal.

The P. E. of a normal group of unselected measures is that distance below or above the median of its surface of distribution whose ordinate cuts off just 25% of the measures between itself and the median. 1 P. E. corresponds by definition to just 25% of the measures next the median. The median is the chosen point of reference, the place from which we begin to count P. E.'s in both directions; it is just zero distance from the median, and corresponds to just 0% of the measures. It is customary to place — P. E. at the left to represent less difficulty, and + P. E. at the right to represent more. By definition 50% of the cases lie between —1

P. E. and +1 P. E.; 25% lie below —1 P. E. and 25% lie above +1 P. E. The area about —4 P. E. holds the very few pupils who solve very few problems; that between the 25 percentile and the 75 percentile holds the great number of pupils who solve an average number of problems; and that about +4 P. E. holds the very few who solve very many problems. To include all the measures, the curve and the base must be infinitely extended. At infinity they meet. Equal numbers and per cents of pupils are represented by equal amounts and per cents of area in the frequency surface, and equal increments of difficulty by equal distances along the base.

We have then the beginning of a table of correspondences between the per cents or tenths of per cents of the measures, and the fractional parts of the unit P. E. Such a table is Table I†. It contains not only the information that 25% taken from the median, corresponds to 1 P. E., and 0% to O P. E.; but also that 0.1% corresponds to .004 P. E., 10.4% to .391 P. E., and so on.

We have now a unit of measure and a table, I, for converting per cents into multiples of that unit.

A few definitions of terms and relations follow.

The *median* is that perpendicular to the base which so divides the surface of frequency that 50% of the measures lie upon either side. Or, the *median* is the mid measure of the group, when the measures have been arranged in the order of magnitude. Or, "The *median*, or 50 percentile, or mid measure, is the place on the scale reached by counting half of the measures in the order of their magnitude, or the place on the scale above which and below which are an equal number of measures." (Thorndike, 1913, page 43).

†Table I is not reproduced here. See Trabue: Completion—Test Language Scales, 1916, page 38; or Woody: Measurements of Some Achievements in Arithmetic, 1916, page 37. In this table per cents are in the vertical column at the left, and tenths of per cents in the horizontal line at the top. The table reads

.000 corresponds to .000 P. E.
.001 corresponds to .004 P. E.
.002 corresponds to .007 P. E.
.100 corresponds to .376 P. E.
.105 corresponds to .395 P. E.
.357 corresponds to 1.582 P. E.
.499 corresponds to 4.600 P. E.

"The *25 percentile* is the measure with three times as many measures above as below it." (Ibid.)

"The *75 percentile* is the measure leaving one third as many measures above as below it." (Ibid.)

The P. E. is the distance upon the base between the median and the median deviation. It corresponds to 25% of the measures.

Two P. E. is twice that distance. If taken between the median and an ordinate 2 P. E. from the median on either side, it corresponds to about 41% of the measures.

Three P. E. is three times that distance. If taken between the median and an ordinate 3 P. E. from the median on either side, it corresponds to about 47% of the measures.

Four P. E. if taken between the median and an ordinate 4 P. E. from the median on either side, corresponds to about 49% of the measures.

Four and six tenths P. E. taken similarly, corresponds to 49.9% of the measures.

The process of deriving the scale will involve the following steps:

(a) Finding the distance between the medians of successive grades. This problem is solved by means of Tables II, III, and IV.*

*The Roman numerals employed in naming the tables have significance as follows: The data obtained from administering the Scientific Method Test to about 500 pupils in each of grades III to VIII are tabulated in two ways, Table II and Table V. These two tables are the crude data from which are derived all the other tables, except Table I, including the final scale. Table II shows the number of pupils getting each consecutive score from 0 to 33 problems right. Table V shows the number in each grade solving each problem. Tables II, III, and IV, form one series, and Tables V, VI, VII, and VIII, form the other. Table IX is built out of the data from both series, and X from IX.

Since there are two basal tables of data, there are two denotations of the word median. In the first series, which deals with scores, median means median score or median number of problems right, or number solved by a *median able pupil.* In the second series, which deals with the number of pupils solving each problem, median means problem solved by the median number of pupils, or problem just hard enough for just 50% of the pupils to solve it, or *median hard problem.* If such a problem existed for the sixth grade, for example, it would be found in Table VI near problem 13, and would be represented there by 50%. Problem 13 is itself very nearly a median hard problem for the sixth grade. It is represented by .495. Problems 18 and 19 are nearly median problems for the seventh grade.

TABLE II

THE NUMBER OF PROBLEMS RIGHT

Grades	III	IV	V	VI	VII	VIII
33						
32						
31					2	
30						1
29				1	1	
28						
27				1	2	3
26					2	6
25			2		6	14
24			2		11	14
23		1	1	5	13	13
22		1	6	7	28	27
21			7	17	33	35
20		2	7	16	35	42
19		3	20	23	43	43
18		4	28	27	27	57
17	1	5	24	29	52	51
16		8	29	29	45	41
15	3	14	29	45	46	37
14	6	17	29	42	35	35
13	6	23	46	39	37	21
12	4	31	42	33	22	15
11	10	27	41	26	14	15
10	9	39	29	21	16	10
9	20	38	33	17	5	9
8	10	42	31	21	7	2
7	29	33	31	14	6	2
6	34	38	17	11	3	4
5	26	36	24	8	1	
4	36	21	15	6	4	1
3	36	31	10	4	2	
2	39	33	8	1		
1	43	23	6	3	1	
0	50	12	1	3	1	
Number tested	362	482	518	449	500	498
Median Scores	4.362	8.334	12.310	14.417	17.096	18.105

(b) Finding the location of zero.

(c) Finding the location of each problem with reference to each grade median. This is accomplished by means of Tables V, VI, VII, and VIII.

(d) Referring all the problems to zero, using the average of their distances therefrom in the several grades. This is accomplished in Tables IX and X.

III. Finding the Distance Between the Medians of Successive School Grades

By this is meant finding the distances MIII–IV, MIV–V, MV–VI, MVI–VII, and MVII–VIII,* or, finding how far it is from the median of each grade to that of the next.

Three steps appear:

1 How many pupils of each grade between the two medians?
2 Each number is what per cent of its group?
3 What P. E. distance does each per cent represent?

Table II shows the distribution of scores in each grade. For instance, 50 out of the 362 third grade pupils scored O; 43 scored 1; 39 scored 2; and 36 scored 4. It shows how the pupils behaved with respect to the problems. It shows how able are the different grades as tested by the specific series of difficulties, and indicates roughly the degree of normality of the different distributions, and the amount of distance between the different medians.

How many pupils of the third grade lie between MIII and MIV? In Table II we read that MIII is 4.362 and MIV is 8.334. The group of 36 third grade pupils who scored 4, includes pupils barely able to solve 4 problems and pupils barely less than able to solve 5. The group, then, extends from 4.00 problems right through 4.99+ problems right. The score 4.362, therefore, falls within group 4, and in the lower half of the group, and the score 8.334 falls in the lower half of group 8. Count .362 of the way into the group scoring 4. .362 of 36 is 13.032 pupils. These belong logically below the median, and leave, of the 36 pupils, 22.968 logically above the median. Before we get to the next median we must add 23, 26, 36, 29, and .334 of the group scoring 8, or .334 of 10. or 3.34, making a total of 115.3. In Table III we read that it is

*The following key will be of use:

M	median
MIII	third grade median
O	zero
O-MIII	interval between O and MIII
MIII-IV	interval between MIII and MIV

TABLE III.

THE NUMBER AND THE PER CENT, WITH P. E. VALUES, OF EACH GRADE LYING BETWEEN ITS MEDIAN AND THE MEDIANS OF THE OTHER GRADES

GRADES	III	IV	V	VI	VII	VIII
III		115.3	162.2	173.5	180.1	
	.318	.448	.479	.4975		
		1.346	2.411	3.015	4.179	
IV	134.4		141.6	193.1	225.5	230.4
	.279		.294	.401	.468	.478
	1.140		1.217	1.909	2.746	2.986
V	228.6	136.7		87.1	164.3	188.9
	.441	.264		.168	.317	.365
	2.318	1.067		0.644	1.340	1.636
VI	211.3	167.5	79.3		101.3	130.3
	.471	.373	.177		.226	.290
	2.811	1.692	0.681		0.891	1.196
VII	244.5	229.7	183.2	116.4		49.8
	.489	.459	.366	.233		.100
	3.395	2.579	1.643	0.922		0.376
VIII		241.3	201.3	155.4	52.1	
		.4845	.404	.312	.105	
		3.200	1.935	1.313	.395	

115.3 pupils, or .318 of the third grade, are between MIIIand MIV, —P. E. value (Table I) 1.346. 130.3 pupils, or .290 of the sixth grade are between MVI and MVIII, P. E., 1.196 The upper number in each group of three is in every case the number of pupils of the grade indicated at the left who lie between the median of that grade and the median of the grade indicated at the top of the column. The second number is the corresponding decimal. The third number is the corresponding P. E.

115.3 pupils from MIII to MIV, or .318 of the 362 measures in the third grade, or 1.346 P. E. This P. E. distance is recorded in Table IV, together with MIV–III, and a number of other intermedian intervals, including several between non-consecutive medians, such as MIV–VIII. By subtraction of shorter from longer intervals we obtain indirect measures of all the intervals. There are also two direct measures of each interval in each case, that upon each of the two distributions concerned. The discrepancy between the various measures of the same distance is so small, and consequently the effect of unequal weighting would be so slight, that all the measures were assigned the weight of 1, and the averages found. Such an average for MIII–IV appears in Table IV. All the intermedian

TABLE IV.
THE DETERMINATION OF THE INTERMEDIAN INTERVALS

3–4	1.346
3–5 – 4–5	1.194
3–6 – 4–6	1.106
3–7 – 4–7	1.433
4–3	1.140
5–3 – 5–4	1.251
6–3 – 6–4	1.119
7–3 – 7–4	.816
AVERAGE	1.176

Measured upon the third grade distribution, the distance MIII–IV is (Table III) 1.346 P. E.; upon the fourth, 1.140 P. E. These are two measures of the same distance, of which six others are given in the same column. MIII–V, measured upon the third grade, minus MIV–V, measured upon the fourth, is a measure of the distance MIII–IV. The number of the grade upon which the measure is taken always occurs first.

intervals have been computed similarly. By addition they become a scale outline composed of nothing but grade medians. They are as follows:

III–IV	IV–V	V–VI	VI–VII	VII–VIII
1.176	1.176	.597	.868	.365

IV. Finding the Location of Zero

In order to locate zero we must proceed with the intervals O–MIII and O–MIV very much as we did with the intermedian intervals. How far is O below the third grade median? That is, how far is it from the 50 third grade pupils who scored zero to the pupil who got the middle score of the grade? The middle score is 4.362 problems. How many, what per cent, and what corresponding P. E. distance lie between the 50 zeros and the 4.362 problems? Add together those getting 1, 2, and 3 right, and enough of those getting 4 right to equal .362 of the 36 in that group, or 13.032 pupils. The total, 131.032 pupils, lies between the score O and the score 4.362. So the third grade median is 131.032 pupils, or .362 of the third grade pupils tested. This .362 corresponds in Table I to 1.616 P. E., which is a measure of the distance O–MIII.

The fourth grade median is, by means of the same process, found to be 2.905 P. E. above O. The other grades show so few zero scores that they are disregarded. 2.905 minus 1.176 is 1.729 P. E., a second measure of O–MIII.

TABLE V.

NUMBER IN EACH GRADE SOLVING EACH PROBLEM

Grades	III	IV	V	VI	VII	VIII
Number Tested	362	482	518	449	500	498
Problems						
1	231	381	454	405	481	489
2	139	265	360	334	445	445
3	115	280	393	361	423	411
5	133	277	379	349	400	405
4	85	232	325	312	392	424
6	107	198	292	298	401	407
15	89	201	270	319	363	373
7	31	129	287	285	388	446
9	46	118	271	300	373	392
14	65	160	282	288	344	335
8	21	98	237	259	377	400
12	48	151	223	249	319	339
17	59	148	228	220	290	289
13	30	106	229	222	318	332
11	33	96	182	207	276	329
10	42	96	182	175	291	327
16	44	95	191	189	273	317
21	38	108	167	189	239	237
18	34	88	146	168	263	260
19	16	38	87	107	183	205
22	9	37	69	93	191	190
24	20	51	81	98	140	186
28	37	85	119	90	117	116
20	29	67	74	78	132	157
33	44	101	111	77	73	57
23	18	35	74	88	111	148
25	2	13	40	51	148	140
27	9	29	62	68	107	105
26	11	27	58	71	99	100
29	8	10	48	44	125	130
30	21	38	71	50	79	102
31	17	34	47	55	62	77
32	13	18	35	4	32	22

231 pupils of the third grade solved problem 1; 139 solved problem 2. The order of problems is approximately that of difficulty.

These two measures are averaged, the first being given double weight because the 50 pupils of the third grade who scored zero seem to give a firmer indication of the location of zero ability than the 12 pupils of the fourth grade who scored zero.

1.616 P. E.
1.616 P. E.
1.729 P. E.

1.654 P. E.

The weighted average distance O–MIII is 1.654 P. E.

By addition of the intermedian intervals with the interval O–MIII, we have a scale composed of a zero point and six medians.

Grade	Distance Of Medians Above O
III	1.654 P. E.
IV	2.830 P. E.
V	4.006 P. E.
VI	4.603 P. E.
VII	5.471 P. E.
VIII	5.836 P. E.

The scale has now to be completed by the addition to it of the numbers representing the difficulties of the 33 problems themselves. To do this means the production of the final scale and involves the two remaining steps.

V. Locating each Problem with Reference to each Grade Median

How far above and below each median is each problem? To answer this question is the same thing as to construct out of the data of each grade a separate scale of the 33 problems. Table VIII shows these six scales. To produce Table VIII, Tables V, VI, and VII were developed. Table V shows the number in each grade solving each problem. For instance, 231 pupils of the third grade and 381 of the fourth solved problem 1. This table can be more easily studied if each entry is changed to a decimal. This

TABLE VI.
THOUSANDTHS IN EACH GRADE SOLVING EACH PROBLEM

	3	4	5	6	7	8
1	638	790	876	902	962	982
2	384	550	695	744	890	894

0.638 of the third grade pupils solved Problem 1; 0.790 of the fourth grade, etc. Figures for Problems 3 to 33 are omitted.

has been done and Table VI is the result. Those problems which show a constant increase, from grade to grade, in the per cents of pupils solving them, are the best for the purpose. Those which behave erratically have to be regarded with suspicion. Obviously, if the sixth grade scores lower upon a given problem than the fifth, that problem does not measure progress from fifth to sixth grade ability. In the earlier editions of the Scientific Method Test a number of such problems were discovered and eliminated.

In order to make a scale of problems each with a stated numerical difficulty or value, some reference point must be adopted. For inches this would be zero distance. For mental ability, zero, instead of being known to begin with, has to be established, and hence is not available as a reference point. Nor is there any absolute point. A point there is, however, which is fixed relatively to the distribution of the measures about it. It is the position which would be occupied by median difficulty for the group, or the middle point of difficulty, or simply the median, — that is, in Table VI, .500. The median occurs as follows in the several grades:

Grades	Between Problems
III	1 and 2
IV	2 and 4
V	9 and 8
VI	12 and 13
VII	21 and 18
VIII	21 and 18

The deviation of each problem from this point is the difference between .500 and the thousandths in each grade solving each problem. It is each entry in Table VI minus .500. It is, for example, .638 minus .500, or .138; .790 minus .500, or .290. In this way

TABLE VII.

DIFFERENCES BETWEEN 0.500 AND THE THOUSANDTHS IN EACH GRADE SOLVING EACH PROBLEM

	3	4	5	6	7	8
1	138	290	376	402	462	482
2	– 116	50	195	244	390	394

Values for Problems 3 to 33 have been found similarly.

Table VII was derived from Table VI. It is the decimal expression of the deviation of each problem from each grade median.

Table VIII is derived from Table VII by substituting for each decimal entry its corresponding P. E. value, read from Table I. Table VIII shows a scale for each grade. The problems range along the base of each distribution in somewhat the same order and with somewhat the same relative difficulties or distances from each other.

VI. Referring all the Problems to Zero According to the Average Position of Each

But we need a single scale for use in all grades alike. Table VIII gives difficulties with reference to each grade median. We have now to change the reference point to zero for each grade, and to state the difficulty of each problem with reference to zero difficulty, and for all the grades in general instead of each in particular. Table IX does this. Problem I is relatively easier for the eighth grade than for any other, and harder for the third, while for all the grades taken together it will average somewhere between. In finding just where it does average, its position in all the grades is considered and each given a weight. Problem 32 in grade VIII, determined as it is by 4.5% of the children of that grade, is not so reliable a measure as problem 22, in the same grade, measured as it is by 38.2%, or as 21, by 47.6%. Nor is problem I (98.2%) very reliable. Per cents near 0% and 100% are to be distrusted and to be weighted low. In averaging the suggestions of each grade as to the position of each problem, those entries in Table VIII within I P. E. of the median are given a weight of 2, those distant I P. E. or more, but less than 3 P. E., a weight of I, while those distant 3 P. E. or more, are omitted. Only three measures are omitted. Problem I is -.523 P. E. from MIII. MIII is 1.654 P. E. from O. Then by subtraction, problem I is 1.131 P. E. from O P. E. This is according

TABLE VIII.

Grades / Problems	III	IV	V	VI	VII	VIII
1	−.523	−1.196	−1.713	−1.918	−2.631	−3.111
2	.437	−.187	−.756	−.972	−1.819	−1.851
3	.706	−.303	−1.038	−1.269	−1.512	−1.391
4	1.076	.071	−.480	−.756	−1.165	−1.549
5	.504	−.280	−.913	−1.130	−1.248	−1.324
6	.799	.334	−.239	−.628	−1.259	−1.346
7	2.026	.918	−.201	−.512	−1.125	−1.867
8	2.331	1.238	.160	−.288	−1.019	−1.269
9	1.692	1.024	−.085	−.644	−.982	−1.186
10	1.772	1.253	.567	.414	−.307	−.600
11	1.979	1.253	.567	.145	−.194	−.616
12	1.649	.723	.261	−.205	−.523	−.698
13	2.054	1.145	.216	.019	−.516	−.640
14	1.357	.644	−.164	−.539	−.727	−.665
15	1.019	.311	−.018	−.825	−.891	−.996
16	1.728	1.264	.500	296	−.172	−.519
17	1.456	.748	.224	.037	−.299	−.303
18	1.953	1.340	.855	.476	−.097	−.085
19	2.530	2.093	1.427	.057	.508	.330
20	2.083	1.609	1.582	1.391	.936	.710
21	1.859	1.125	.685	1. .296	.082	.089
22	2.905	2.114	1.69	1.211	.445	.445
23	2.439	2.155	1.582	1.269	1.135	.786
24	2.370	1.851	1.449	1.155	.864	.476
25	3.725	2.857	2.114	1.795	.795	.855
26	2.789	2.357	1.803	1.487	1.259	1.243
27	2.905	2.305	1.742	1.524	1.176	1.191
28	1.884	1.380	1.096	1.248	1.076	1.081
29	2.986	3.015	1.962	1.918	1.000	.945
30	2.331	2.093	1.622	1.811	1.487	1.222
31	2.483	2.177	1.979	1.720	1.713	1.506
32	2.2667	2.631	2.211	3.506	2.257	2.514
33	1.728	1.196	1.176	1.409	1.563	1.780

P. E. Deviations of each problem from each median, or P. E. Equivalents of the dfferences between 0.500 and the Thousandths in each grade solving each problem.

to the third grade distribution. According to the fourth, problem I is 1.634 above O P. E.; the fifth, 2.293; the sixth, 2.685; the seventh, 2.840. These results are recorded in Table IX, which gives the

TABLE IX.

	1	2
	1.131	2.091
	1.131	2.091
	1.634	2.643
	2.293	2.643
	2.685	3.250
	2.840	3.250
		3.631
		3.631
		3.652
		3.985
Averages	1.952	3.087

Table showing the distance above zero of Problems 1 and 2 as indicated by each grade, and also as indicated by all the grades averaged together. Problem 1 is 1.131 P. E. above zero as indicated by the third grade. To this indication double weight is given, as explained in the text. Problem 1 is 1.634 P. E. above zero as indicated by the fourth grade; 2.293 by the fifth, etc. In the eighth grade it is more than 3 P. E. from the median, and is therefore disregarded. The weighted average of the various indications is, for problem 1 1.952 P. E., and for problem 2, 3.087 P. E. Problems 3 to 33 have been evaluated in the same manner. The values appear in Table X, Column B.

weighted average position of each problem above zero. Column B in Table X is these distances or P. E. values arranged in the order of their magnitude. This is the scale. The entries in column C, however, are sufficiently accurate and more convenient.

TABLE X.

A	B	C
1	1.952	20
3	3.060	31
2	3.087	31
5	3.090	31
4	3.541	35
6	3.542	35
7	3.920	39
15	3.939	39
9	4.065	41
14	4.146	41
8	4.254	43
12	4.356	44
17	4.492	45
13	4.567	46
11	4.735	47
10	4.749	47
16	4.752	48
21	4.960	50
18	4.991	50
28	5.361	54
33	5.542	55
19	5.561	56
24	5.658	57
22	5.671	57
20	5.708	57
23	5.770	58
30	5.828	58
27	5.874	59
26	5.890	59
31	5.996	60
29	6.194	62
25	6.303	63
32	6.415	64
		1582 Total

A, problem numbers. B, scale of values of the problems, carried to nearest thousandth. C, scale, carried to nearest tenth of a P. E. and multiplied throughout by 10.

AGE-GRADE DISTRIBUTION, ITS PRINCIPLES AND APPLICATION

ERNEST C. WITHAM

Superintendent of Schools, Southington Ct.

THERE are two chief factors in the consideration of the efficiency of any school system. 1. ***Quality of Instruction.*** This is evaluated best by means of the various standard tests. 2. ***Age and Grade Distribution of the Pupils.*** Either one of these factors without the other does not tell the story, because we may have a school system which makes a fine showing on all the tests, and yet the retardation may be very great, and the percentage of over-age pupils (pupils too old for the grade they are in) may be very large. There is surely no virtue in a system that takes 2 to 4 years more than is necessary to do its allotted work. On the other hand, a school system that always promote everyone, and in which the percentage of over-age pupils is very small, but where the pupils' work is of inferior quality, such a system would not be efficient, because the small over-age percentage would mean nothing, except that pupils were being promoted regardless of what they know or can do.

It is necessary first of all to have a clear understanding of the terms used in an investigation of this kind.

With regard to age-grade distribution pupils are divided into three groups: 1. Below Normal, 2. Normal, 3. Above Normal.

Below Normal–these pupils too old for their grade.
Normal—these pupils of right age for their grade.
Above Normal–these pupils young for their grade.

The normal ages for grade 1 are 6-7 years, for grade 2, 7-8 years and so on for the remaining grades. All pupils older than these ages are below normal, and all younger than these ages are above normal. These are the most commonly accepted limits of age-grade distributions.

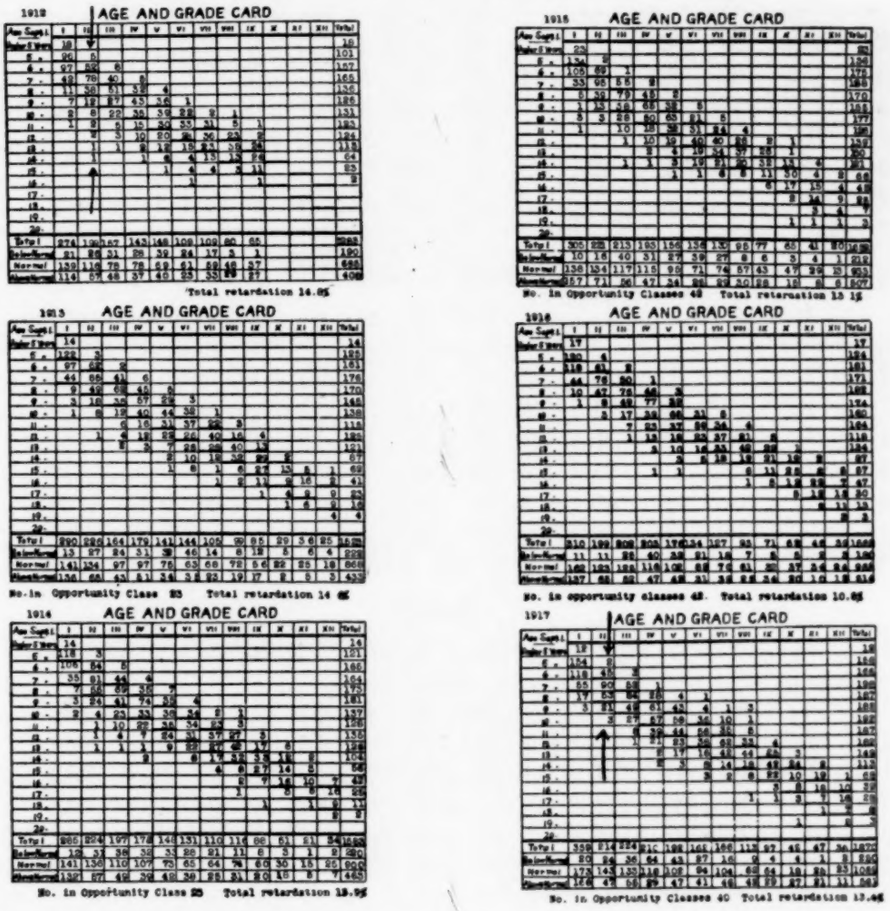
The present investigation covers a period of six years in a growing school system of from 1500 to 2000 children. The information has been tabulated each year by schools and then assembled using the following form:

AGE AND GRADE CARD

Age Sept. 1.	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	Total
Under 5 Years													
5 "													
6 "													
7 "													
8 "													
9 "													
10 "													
11 "													
12 "													
13 "													
14 "													
15 "													
16 "													
17 "													
18 "													
19 "													
20 "													
Total													
Below Normal													
Normal													
Above Normal													

The age-grade data is always determined the first month of the school year, and the pupils' ages are always figured from September 1. The heavy serrated lines on the blank separate the three groups, the normal group coming between them and the below normal group, below and the above normal group, above the serrated lines.

AGE GRADE DISTRIBUTION FOR SIX YEARS


1912 AGE AND GRADE CARD
Total retardation 14.0%
1913 AGE AND GRADE CARD
No. in Opportunity Class 23 Total retardation 14 0%
1914 AGE AND GRADE CARD
No. in Opportunity Class 25 Total retardation 13.9%
1915 AGE AND GRADE CARD
No. in Opportunity Classes 48 Total retardation 13 1%
1916 AGE AND GRADE CARD
No. in opportunity classes 45. Total retardation 10.0%
1917 AGE AND GRADE CARD
No. in Opportunity Classes 40 Total retardation 13.0%

In 1912 it will be noticed that in the second grade there were pupils all the way from 5 years old to 14 years old. Of course such a condition is bad, but was unavoidable until opportunity classes were organized, which now take care of the worst cases of over-ageness. In 1917 the span of ages for the second grade is 5 to 10, a much better showing.

The percentages of over-age or below normal pupils has apparently not been large for several years at least, yet there has been some improvement from year to year, as a result of careful study of the situation.

The following tables are taken from the age-grade distribution for the past six years.

Above Normal

	1	2	3	4	5	6	7	8	9	10	11	12	Total	Per Cent
1912	114	57	48	37	40	23	33	29	27				408	31.8
1913	136	65	43	51	34	35	23	19	17	2	5	3	433	28.4
1914	132	57	49	39	42	38	25	31	20	18	5	7	463	29.2
1915	157	71	56	47	34	26	29	30	28	15	8	6	507	30.7
1916	137	65	52	47	42	31	39	25	34	20	10	12	514	30.9
1917	166	47	55	29	47	41	46	42	29	27	21	11	561	30.

Below Normal

	1	2	3	4	5	6	7	8	9	10	11	12	Total	Per Cent
1912	21	26	31	28	39	24	17	3	1				190	14.8
1913	13	27	24	31	32	46	14	8	12	5	6	4	222	14.6
1914	12	31	38	32	33	28	21	11	8	3	1	2	220	13.9
1915	10	16	40	31	27	39	27	8	6	3	4	1	212	13.1
1916	11	11	25	40	32	21	18	7	5	5	2	3	180	10.8
1917	20	24	36	64	43	27	16	9	4	4	1	2	250	13.4

Normal

	1	2	3	4	5	6	7	8	9	10	11	12	Total	Per Cent
1912	139	116	78	78	69	61	59	48	37				685	53.4
1913	141	134	97	97	75	63	68	72	56	22	25	18	868	57.
1914	141	136	110	107	73	65	64	74	60	30	15	25	900	56.9
1915	138	134	117	115	95	71	74	57	43	47	29	13	933	56.5
1916	162	123	125	116	102	82	70	61	32	37	34	24	968	58.3
1917	173	143	133	118	102	94	104	62	64	18	25	23	1059	56.6

Two methods of attack have been used in studying the above data. 1. Percent of all in each group by grades. 2. Percent in the three groups by grades.

First Method

In the first method we have 100% below normal, 100% normal and 100% above normal. The percent of all the first group in each

grade has been worked out, that is, 11.1% of all the below normal pupils are in the first grade, 13.7% are in the second grade, etc. The following tables give the percentages for the three groups, derived from the first three tables.

Per Cent of the "Below Normals" in Each Grade

Year	Grades											
	1	2	3	4	5	6	7	8	9	10	11	12
1912	11.1	13.7	16.3	14.7	20.5	12.6	8.9	1.6	.5			
1913	5.9	12.2	10.8	14.0	14.4	20.7	6.3	3.6	5.5	2.2	2.7	1.8
1914	5.5	14.1	17.3	14.5	15.0	12.7	9.5	5.0	3.6	1.4	.5	.9
1915	4.7	7.5	18.9	14.6	12.7	18.4	12.7	3.8	2.8	1.4	1.9	.5
1916	6.1	6.1	13.9	22.2	17.8	11.7	10.0	3.9	2.8	2.8	1.1	1.7
1917	8.0	9.6	14.4	25.6	17.2	10.8	6.4	3.6	1.6	1.6	.4	.8
Ave.	6.9	10.5	15.3	17.6	16.3	14.5	9.0	3.6	2.8	1.9	1.3	1.1

Per Cent of the "Normals" in Each Grade

Year	Grades											
	1	2	3	4	5	6	7	8	9	10	11	12
1912	20.2	16.9	11.4	11.4	10.1	8.9	8.6	7.2	5.4			
1913	16.2	15.4	11.2	11.2	8.6	7.3	7.8	8.3	6.5	2.5	2.9	2.1
1914	15.7	15.1	12.2	11.9	8.1	7.2	7.1	8.2	6.7	3.3	1.7	2.7
1915	14.8	14.4	12.5	12.3	10.2	7.6	7.9	6.1	4.6	5.0	3.1	1.4
1916	16.7	12.7	12.9	12.0	10.5	8.5	7.2	6.3	3.3	3.8	3.5	2.5
1917	16.3	13.5	12.6	11.1	9.6	8.9	9.8	5.9	6.0	1.7	2.4	2.2
Ave.	16.6	14.7	12.1	11.6	9.5	8.1	8.1	7.0	5.4	3.2	2.7	2.2

Per Cent of the "Above Normals" in Each Grade

Year	Grades											
	1	2	3	4	5	6	7	8	9	10	11	12
1912	27.9	14.0	11.8	9.1	9.8	5.6	8.1	7.1	6.6			
1913	31.4	15.0	9.9	11.8	7.9	8.1	5.3	4.4	3.9	.5	1.2	.7
1914	28.5	12.3	10.6	8.4	9.1	8.2	5.4	6.7	4.3	4.0	1.1	1.5
1915	31.0	14.0	11.0	9.3	6.7	5.1	5.7	5.9	5.5	3.0	1.6	1.2
1916	26.7	12.6	10.1	9.1	8.2	6.0	7.6	4.9	6.6	3.9	1.9	2.3
1917	29.4	8.4	9.8	5.2	8.4	7.3	8.2	7.5	5.2	5.0	3.7	1.9
Ave.	29.1	12.7	10.5	8.8	8.3	6.7	6.7	6.1	5.3	3.3	1.9	1.5

Figure 1, is a graph of the below normal group of percentages for the six years. The similarity of the curves for the various years should be noted.

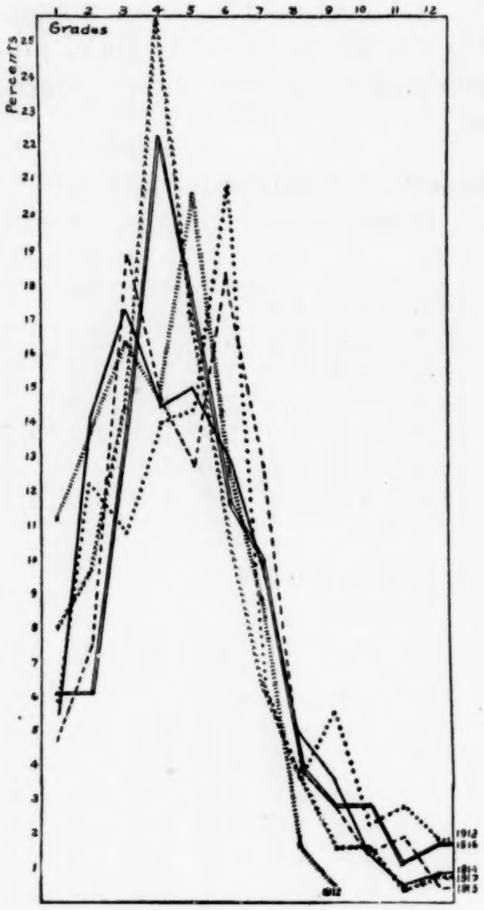


FIGURE 1.

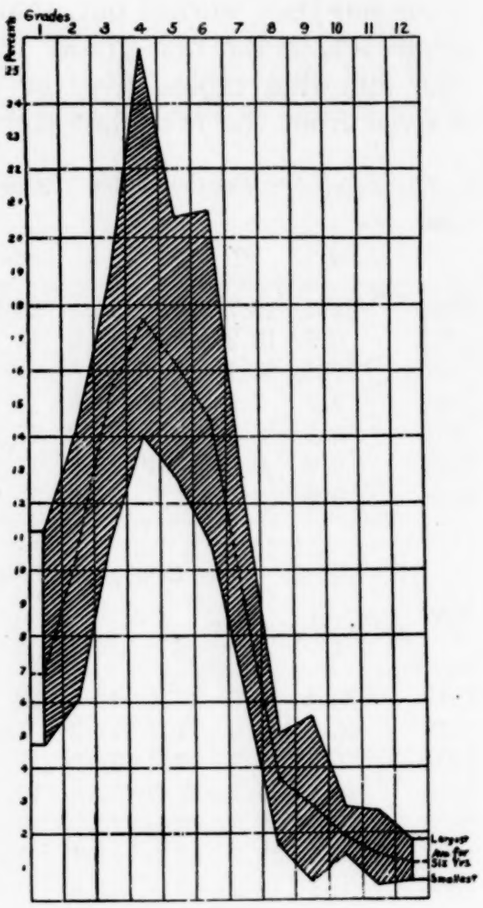


FIGURE 2.

In figure 2 is shown the largest and smallest percentage of below normal pupils for each grade for the six years, and also the dotted line in the center shows the average percentage of below normal pupils for the six year period. This is probably an accurate picture of the condition. It will be noted that the greatest percentages of below normal pupils are in the 3, 4, 5 and 6 grades, but especially in the 4 grade.

What is the reason for so much more retardation in the fourth grade than in any other grade? This is a fair question, and there are probably several answers to it. This is worth thinking about. How do the fourth grades usually come out on the standard tests? All of this is given not to tear down or cast suspicion in any way, but rather in the spirit of constructive criticism, studying the system,

and improving the weakest spots. Is the course of study too difficult for fourth grades? There are several questions that will occur to all concerned at this point. It is our job to solve the problem, the same one that other places are facing if they have yet got to the stage of investigation.

It should be noted that the early age at which pupils are allowed to first enter school, precludes the possibility of much over-ageness in the first two grades.

The school board requires pupils under 16 to finish the sixth grade before leaving school, so that it is easy to see why the intermediate grades have the greatest amount of the below normals. But look at figure 2 and see how many there are in the sixth grade and what a sudden drop there is then. What does this mean? It says very plainly that nearly all pupils who are greatly retarded in the upper grades, leave school just as soon as the law will allow them to. Here are six years of proof that it does not pay to make old pupils repeat the intermediate grades. It simply means driving them out of school at the first opportunity they have for leaving.

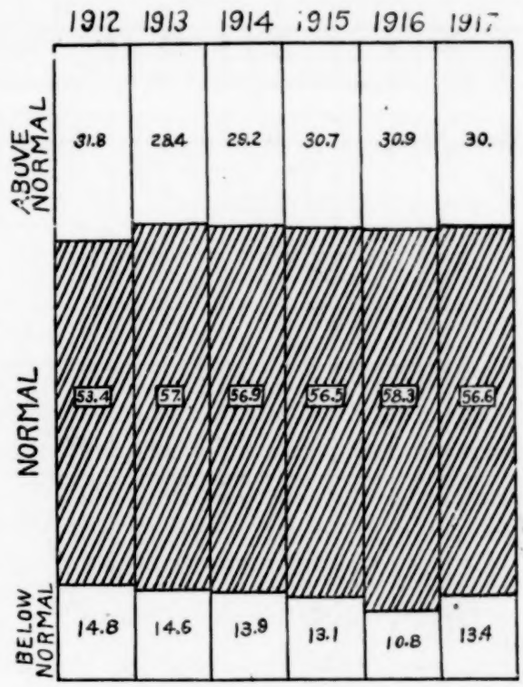


FIGURE 3.

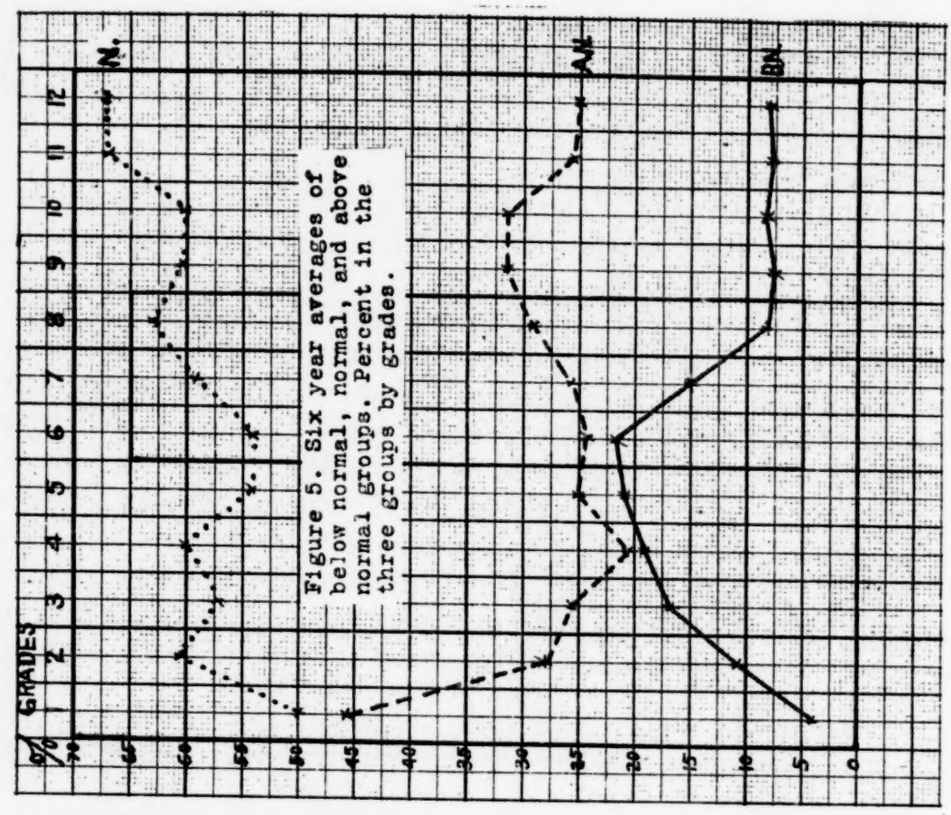


Figure 5. Six year averages of below normal, normal, and above normal groups. Percent in the three groups by grades.

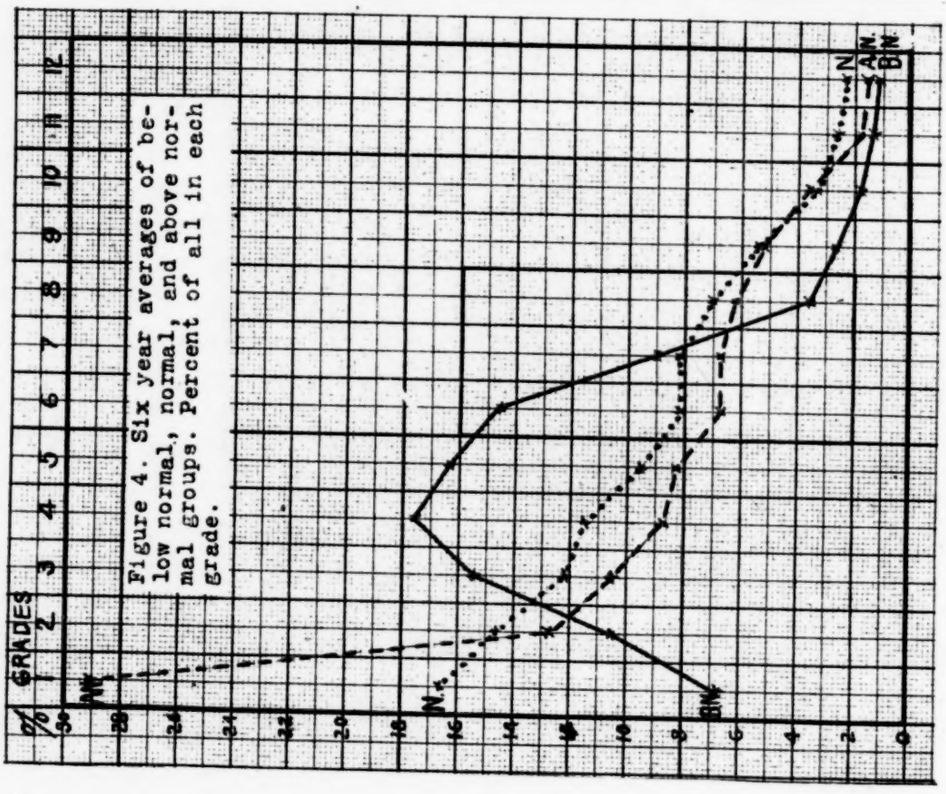


Figure 4. Six year averages of below normal, normal, and above normal groups. Percent of all in each grade.

Figure 3 gives both graphically and numerically the percentages of normal, below normal and above normal pupils for the past six years.

Figure 4, gives the graphs of the six year averages of the three groups in percent of all in each group by grades.

The peculiarities of the first grade are to be explained by the following facts:

1. Early admission (If pupils will be 5 during fall term they are permitted to begin school in September).
2. No Kindergartens
3. Large numbers are from non-English speaking homes, so that it takes two years for many to do the first year work.

These facts all tend to make the first grade very large and also to make the above normal group very large.

A comparison of 6-7-8 grade pupils in the three groups show clearly what has already been said, that the Below Normal group leave school while the other two groups of pupils remain in school during these years.

Second Method

The second method used in this study finds out the percent of below normal, normal and above normal pupils in each grade.

Per Cent of pupils in the three groups by grades

Grades	1	2	3	4	5	6	7	8	9	10	11	12
Below Normal	4	11	17	19	21	22	15	8	7	8	7	8
Normal	50	61	57	60	54	54	59	63	61	60	67	67
Above Normal	46	28	26	21	25	24	26	29	32	32	26	25

Figure 5, graphically shows the comparison of three groups. Notice the below normal graph in this figure is very similar to that in figure 4. This shows conclusively that whatever way the below normal data is handled, we get the same general result. Also the comparison of 6-7-8 grades shows even more clearly than in the first method what happens to the three groups.

Comparison with other Places

A large number of school reports have been examined with the view of securing reliable data for comparison with other places. While perhaps 50% of school systems now publish age-grade dis-

tributions, this seems to be done in such a perfunctory way that it is very difficult to find much help in school reports.

In the first place the nomenclature is very different. Some employ the terms Young, Normal and Old; others Accelerated, Normal and Retarded; Under Normal Age, Normal Age and Over Normal Age; and Below Normal, Normal and Above Normal are also used in about the same proportion. Most all use Below Normal to mean the pupils too old for their grade, but there are numerous instances where it is used for the pupils young for their grade.

There is also diversity of practice in the dates employed for securing the data. September 1 and January 1 seem to be the most common now in use.

Practice is about evenly divided in the matter of the arrangements of tabulating the data. Some use the form with the grades horizontally and the ages vertically and others with the opposite arrangements. If we all used the same form we could easily compare results, as it is now comparison is difficult.

Another difference of opinion and practice is in regard to the age limits of the three groups. The majority use 6-7 for the normal ages of the first grade, but 6-6½ and 5-6 and other combinations are frequently found.

Some tabulate age-grade data by straight grades and others by 1a, 1b, 2a and 2b, etc. This would not be so bad but for the fact that in some places an A division stands higher than a B division, while in other places B divisions are higher than A divisions. Sometimes the distribution is made separately for the boys and girls.

After the age-grade data has been secured there is not much evidence that it is analyzed and put to any practical use. Most superintendents seem to stop with just the distribution table. Some few work out the percent of one group, seldom two groups and almost never does anyone work out the percentages of the three groups, but when they do, the second method is the one used. Very seldom is there any comment with the age-grade table. It looks as though it was mechanically assembled because it is in style to have such a table in one's school report, but the facts are apparently not studied, or comparisons made. Even Bliss' excellent recent book on "Methods and Standards for Local School Surveys" does not suggest means of securing age-grade data in a systematic and common way. On page 47 of this book there is a list of about 30 cities giving the

"Number of pupils and Percentage Classification" in age and progress. The facts here presented are of doubtful comparative value because of the differences practiced in procuring the original data.

Information of a reliable nature has been secured for three other cities and is here given.

Percentages of Distribution

	1	2	3	4	5	6	7	8	9	10	11	12	T
1.Hackensack N. J. (Jan. 1917)													
Below Normal	13	13	10	12	12	9	7	14	14				11
Normal	78	78	72	63	54	58	67	65	65				67
Above Normal	9	9	18	25	34	33	26	21	21				22
2. Waterbury Ct. (Sept. 1917)													
Below Normal	8	12	15	20	32	30	17	13	7	7	8	8	18
Normal	88	57	61	56	50	50	62	64	61	55	67	62	60
Above Normal	4	31	24	24	18	20	21	23	32	38	25	30	22
3. Norwich Ct. (Sept. 1915)													
Below Normal	8	9	14	23	27	28	14	13					16
Normal	57	61	64	57	58	54	64	71					59
Above Normal	35	30	22	20	15	18	22	16					25

1. From Annual Report of Public Schools 1917
2. From Annual Report of the Board of Education 1917
3. From Annual Report of Public Schools 1916

The number of cases or records involved in the three cities together with the six years in Southington gives nearly 30,000 cases.

Grades	1	2	3	4	5	6	7	8	9	10	11	12	T
Percent Below Normal													
Hackensack	13	13	10	12	12	9	7	14	14				11
Waterbury	8	12	15	20	32	30	17	13	7	7	8	8	18
Norwich	8	9	14	23	27	28	14	13					16
Southington	4	11	17	19	21	22	15	8	7	8	7	8	13
Ave.	8.3	11.3	16	18.5	23	22.3	13.3	12	9.3	7.5	7.5	8	14.5
Percent Normal													
Hackensack	78	78	72	63	54	58	67	65	65				67
Waterbury	88	57	61	56	50	50	62	64	61	55	67	62	60
Norwich	57	61	64	57	58	54	64	71					59
Southington	50	61	57	60	54	54	59	63	61	60	67	67	57
Ave.	68.3	64.3	63.5	59	54	53	63	65.8	62.3	57.5	67	64.5	60.8
Percent Above Normal													
Hackensack	9	9	18	25	34	33	26	21	21				22
Waterbury	4	31	24	24	18	20	21	23	32	38	25	30	22
Norwich	35	30	22	20	15	18	22	16					25
Southington	46	28	26	21	25	24	26	29	32	32	26	25	30
Ave.	23.5	24.5	22.5	22.5	23	23.8	23.8	22.3	28.3	35	25.5	27.5	24.7

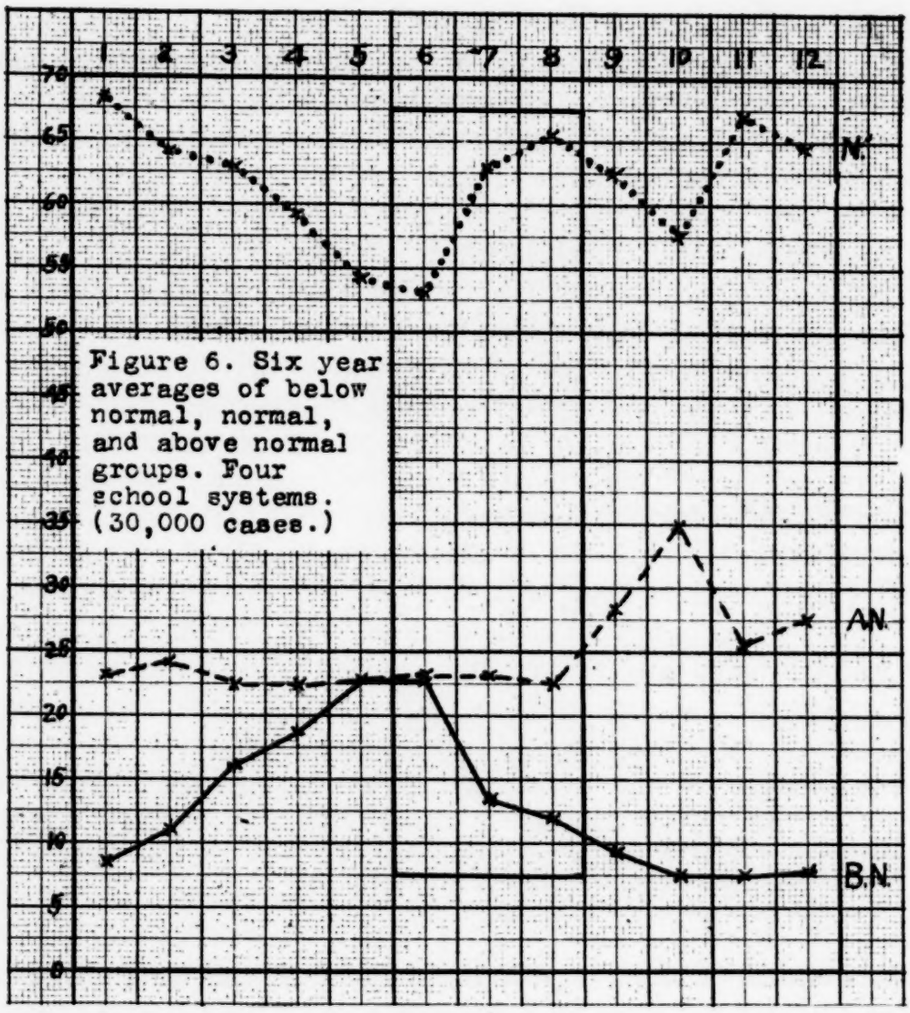


Figure 6, gives the percentage age-grade distribution of the three groups for the four school systems. The below normal graph is very similar to the one in figure 5. Again a comparison of 6-7-8 grades shows that normal and above normal age pupils remain in school while nearly all of the over-age pupils leave school. Some of the latter class return for a few weeks in the seventh grade and so are credited as being in that grade, but they soon nearly all drop out.

VARIABILITY IN MEMORY SPAN*

DAVID MITCHELL

New York City

THE memory span of individuals of the same age has been known to vary in quite marked degree. A person of very low grade intelligence may have a memory span of 3, 2, or 1 or may even be unable to reproduce one impression. On the other hand exceptional individuals have been able to reproduce 10, 11, 12 or even a greater number of discrete impressions. In the same way they vary in every known function. Measurements of ability to read show that in the same age group there are wide variations. These variations are such that in the ten year group there are children who are not able to read any better than some children in the nine year group. Some children of the eight year group excel some of the twelve year group. In the different age groups there is overlapping, the children of one age varying so much that always some of the one age are better than some of those, one, two, three, or four years older.

In a large measure we have tended to ignore this variability and to think that the average performance was typical of every child. The child of eight years should be able to spell certain words, should be able to do certain mathematical operations and should be able to read paragraphs containing words of more or less well known difficulty. All children have been supposed to reach and none have been supposed to greatly exceed this average performance ability. Only recently has this status of fixed averages for age groups been seriously questioned.

The variability in an individual's performance has to a much greater extent been overlooked. Increase in performance ability, corresponding with age increase, has been recognized, but that a child should be better in a performance one day than he would be in a similar performance on the following day has not been generally conceded. If a child is able to read a certain passage on Monday, he should be able to read it as well on Tuesday, even if he has had no practice in the interval, and should read it better if an opportunity for practice has been provided. Furthermore, if he has been able to spell a certain group of words the first day of the month, we

*Paper read at The New York Branch of the American Psychological Association, November, 1918.

have assumed that he should be able to spell those words with the same degree of accuracy at the end of the month. If he has been able to repeat correctly the multiplication table on Friday, his ability to do this on the following Monday has not been questioned. Our norms of performance ability have been established on the basis of a belief in the constancy of the individual's performance.

The constancy of average results has confirmed us in our notion of the individual's constancy. The repeated testing of groups of individuals has given average results and measures of deviation which were approximately identical from day to day and from group to group. A study of the individual child and his status from day to day in the group might have shaken our faith in the reliability of individual results, but so far little study of this kind has been made. In order to evaluate the assumption of the constancy of an individual's performance, it was decided to test one function which is supposed to be directly dependent upon neurone excitability. This function cannot be materially changed by practice although it does improve with age. The reproduction of discrete elements, being considered a function of the organic memory, should show as little variation as a function such as mathematical calculation or spelling a certain number of words, since performance in mathematical calculation and in spelling depends largely upon the training or practice which one has had.

Subjects

The children chosen for the experiment were included in five fifth grade groups in one school and three seventh grade groups in other schools. In the five fifth grade classes there were over two hundred children. In the three seventh grade classes there were over one hundred and twenty-five children. Of the fifth grade children, 192 were given a Memory Span test for digits on two days which were separated by a period of exactly five weeks. Of the seventh grade children, 97 were given the Memory Span test on ten successive school days. The fifth grade group included 103 girls and 89 boys; the seventh grade classes 53 boys and 44 girls.

Procedure

The series of digits for each of the trials were made up by a chance selection of numbers, care being taken to eliminate sequences either forward or backward. The presentation of the numbers was given

orally to the group in the various classes and the reproductions were written by the pupils. As the experimenter entered the room at the time of a regular class session, the teacher distributed the record blanks and asked the children to clear their desks and to be certain to have pencils ready. The experimenter then gave instructions in the following way: "We are going to do an experiment this morning. In an experiment we try to find out something. I am going to try to find out how many numbers you can remember. In an experiment it is important that everyone should do the same thing in the same way, so I am going to ask you to pay very close attention to what I tell you and then do exactly as I say. First I want you to write your name at the top of this sheet of paper which has been given to you. Then write your age, your grade, the date, and whether you are a boy or girl." When the children had completed these instructions the experimenter proceeded. "I am going to say some numbers to you and after I have finished saying them I want you to write them down. For instance, if I should say 2-4-6, you would write down at the top of the paper 246 (indicating the method by writing the numbers on the board). Then if I should say 5-7-9, you would write 579 under the other number in this way (indicating them on the board in the right position). You would do this for as many numbers as I might give you. Do you understand? It is important that no one should start to write until I have given all the numbers, so I am going to ask you to hold up your pencils like this (the experimenter holds his pencil up at the level of his shoulder). I will also hold my pencil up and when I bring it down like this (indicating by the movement of the hand to the level of the desk), that will be a signal to you to write down the numbers I have just given you. As soon as you have finished writing them down I want you to hold up your pencil again and that will be a signal to me that you are ready for the next series of numbers. Do you all understand? Then let's begin. Everyone put his pencil up. First I will give you three numbers. Ready!" The test begun, the succeeding series of numbers were given as soon as all the pencils had been raised. The numbers were dictated as distinctly as possible in a monotonous tone of voice at the rate of one per second. The series given were two trials on each series of from 3 to 10 digits.

A method of scoring the Memory Span was a matter of serious question. Consideration was given to the method of scoring suggested by the modified Spearman's Foot rule Correlation $r = 1 - \frac{\Sigma\ d.^1}{(n^2-1)/3}$ The amount of labor involved prohibited the use of this method. We therefore, adopted an arbitrary ruling. Two scores were made. The first score was the longest series, in both attempts at which the child was successful at reproduction and in no series previous to which had he failed. This score we may call the "Consecutive Series" score. The other or "Maximum Series" score is the longest series which the child reproduced, regardless of the number on which he had previously failed. A third method of scoring might have been adopted which would give as the child's score the longest series on which the child correctly reproduced one of the trials, and on no series previous to which had he failed on both the trials. Since the question was the variability and not the determination of an absolute standard for different groups the two scores mentioned seemed sufficient.

Results

Tables I and II present the "Consecutive Series" score and the "Maximum Series" score respectively for the 192 children who were tested on two days separated by the period of five weeks.

TABLE I

"Consecutive Series" Score

Trial I \ II	3	4	5	6	7	8	9	10	Total
3	2	2	1						5
4	5	16	23	5					49
5	5	20	26	13					64
6	2	6	21	21	10	1		1	62
7			2	6	3				11
8		1							1
9									
10									
Total	14	45	73	45	13	1	1		192

[1]WHIPPLE, G. M. *Manual of Mental and Physical Tests.* Vol. II, p. 161.

TABLE II

"Maximum Series" Score

Trial I \ II	3	4	5	6	7	8	9	10	Total
3									
4		2	2	2					6
5		1	13	18	3		1		36
6		2	17	40	18	8	4	1	90
7			2	7	9	4	4		26
8				1	6	3	1		11
9			1	2		1	1	2	7
10		1	2	3	3	1	2	4	16
Total		6	37	73	39	17	13	7	192

These are correlation tables showing the number of children who made the various scores on the two trials. The scores for trial I are shown in the first column. The scores for trial II are shown in the first line of the table, beginning with the second column. In the table of "Consecutive Series" scores we find that two children had a score of three on each of the two trials; two children scored three on the first trial and four on the second trial; one child whose score was three on the first trial scored five on the second trial. In the third line and third column of Table II we find that two children had a score of four in the "Maximum Series" on the first trial and four on the second trial. The next column of the same line shows that two children scored four and five on the first and second trials respectively. Two children also scored four and six in the first and second trial respectively. In both tables the number of children who had the same score on the two trials is underlined.

The number of children who gave the same score on the two trials, or whose score differed by one, two, or three or more numbers is shown in Table III.

TABLE III

Relation of Scores on Two Trials

"Consecutive Series" Score		"Maximum Series" Score
68	Equal Score	72
100	Different by One	79
20	Different by Two	23
2	Different by Three	9
2	Different by Four	6
0	Different by Five	2
0	Different by Six	1

The first column is for the "Consecutive Series" score. It shows that 68 children had the same score on the two trials. One hundred children scored one more in one trial than they did in the other. Twenty children had scores on the two trials different by two while four children's scores differed by three and four. Similar results are presented in the final column for the "Maximum Series" score. By this method of scoring the variation in the two trials is greater. There were only 151 children whose score was within one of being the same for the two trials. Of the other children 18 varied three or more, two of them varying as much as five and one as much as six.

Table IV shows the record of all children with the averages and medians for both methods of procedure and each trial. The first column of the table gives the scores. The second and third columns contain the numbers of children making the various "Consecutive Series" scores on the two trials.

TABLE IV

Number of Children Making Different Scores

	"Consecutive Series" Score		"Maximum Series" Score	
	Trial		Trial	
Score	1st	2nd	1st	2nd
3	5	14	0	0
4	49	45	6	6
5	64	73	36	37
6	62	45	90	73
7	11	13	26	39
8	1	1	11	17
9	0	0	7	13
10	0	1	16	7
Average	5.1	5.0	6.4	6.5
Median	5	5	6	6

Columns four and five contain the corresponding "Maximum Series" score. The numbers in columns two and three are not identical nor are the numbers in columns four and five. They are sufficiently alike to make the average and the median almost the same for the two trials. The median for the "Consecutive Series" score is the same in both trials and the average differs by only 0.1.

In the "Maximum Series" score the median is again the same, the average score in the two trials differing by only 0.1. The average or group performance for the two days is practically identical. We are not, however, justified in saying that the children show a constancy of performance since they have been shown in Table III to vary considerably.

Tables V and VI show results for the 97 children who were tested on ten successive days.

TABLE V.

"Consecutive Series" Score.

Average Scores and Number of Children Making Them.		Actual Scores on the Different Trials									
		3	4	5	6	7	8	9	10	0	Total
3.0											
3.5	1	1	6	1	1					1	10
4.0	4	4	24	7	4					1	40
4.5	10	5	36	37	21					1	100
5.0	23	5	52	77	82	12	1			1	230
5.5	26	2	32	48	127	41	8			2	260
6.0	16		13	14	80	40	13				160
6.5	5		4	2	15	19	9	1			50
7.0	9		1	3	20	29	30	7			90
7.5	3		2		3	9	7	6	3		30
8.0											
8.5											
9.0											
9.5											
10.0											
Total		17	170	189	353	150	68	14	3	6	970

In the first part of the first column, the average scores in nearest half units are shown:—that is, if a child made a score of four on six of the ten days, a score of three on two other days, a score of five on the remaining two days, his average score for the ten days would be four. Four children, whose average score individually was four, had scores similar to these. In the second part of the column, the number of children who made the different average scores is shown. In the first line of the table the different scores are shown. For example, in the line for the average score of 7.5, it is shown that three children made this average. Two of the 30 trials gave scores of four, three of six, nine of seven, seven of eight, six of nine, and

three of the trials gave scores of ten. The last line in the column gives the number of times each of the different scores was made in the 970 trials made by these 97 children.

The construction of Table VI is identical, this table being for the "Maximum Series" score.

TABLE VI.

"Maximum Series" Score

Actual Score on the Different Trials

Average Score and Number of Children Making Them.	3	4	5	6	7	8	9	10	0	Total
3.0										
3.5										
4.0										
4.5										
5.0	2	3	8	8	1					20
5.5	1	1	2	7						10
6.0	12	2	14	61	37	6				120
6.5	20		11	84	71	26	7	1		200
7.0	17		2	38	74	43	11	2		170
7.5	17			17	47	81	23	2		170
8.0	9			5	13	37	27	8		90
8.5	9				6	30	37	17		90
9.0	7					15	31	24		70
9.5	3					1	10	19		30
10.0										
Total		6	37	220	249	239	146	73		970

As the average "Maximum Series" score increases, we see that the numbers are sharply cut off in the last column of the table. It was assumed in the beginning of the experiment that very few children would correctly reproduce a series of ten digits. We find that 73 out of the 970 trials result in a "Maximum Series" score of 10. It is obvious from these results that the series should have been increased in length to at least twelve, and possibly a higher number.

The question is raised as to the average daily score in these ten days. Table VII presents the result.

TABLE VII.

Day	Average Daily Scores "Consecutive Series" Score	"Maximum Series" Score
1	5.5	6.9
2	5.8	7.2
3	4.8	7.3
4	5.6	7.2
5	5.8	7.4
6	6.0	7.6
7	6.0	7.8
8	5.7	7.6
9	5.9	7.6
10	6.0	7.8

In the first column of this table the day on which the trial was given is shown. The second column gives the average "Consective Series" score for the days, and the third column gives the average "Maximum Series" score. In the first case the range of scores is from 4.8 to 6.0, a range of 1.2. This range is wide and should be kept in mind when considering the later discussion of correlations in which the results of trial III are correlated with results from each of the other trials. The reason for this lower score which is 0.8 lower than any other score in the column is found in the fact that the experimenter for one group of children changed his method of procedure for that day. He consciously lengthened the time between the enunciation of the digits in a series. The result was quite marked;—not one of the children with whom he worked succeeded in getting a "Consecutive Series" score higher than four on that day.

The range of variability in averages with the "Maximum Series" score is somewhat less. The lowest score is 6.9 and the highest is 7.8, a range of 0.9. The group results are similar from day to day as was the case with the 192 children whose averages for the "Consecutive Series" score differed by only 0.1 and the averages of the "Maximum Series" score showed only the same difference. Group results are approximately identical from day to day but individual scores vary considerably.

In order to get a simpler indicator of individual variabilities coefficients of correlation were calculated according to the formula

$$r = \frac{p^{2}}{\sigma \times \sigma y}$$

[1]See page 181ff. *Introduction to the Theory of Statistics*, by G. UDNY YULE.

TABLE VIII.

Trials	Coefficients of Correlation	
	"Consecutive Series" Scores	"Maximum Series" Scores
I-II	0.40	0.54
I-III	0.09	0.49
I-IV	0.46	0.50
I--V	0.30	0.59
I-VI	0.19	0.64
I-VII	0.31	0.59
I-VIII	0.31	0.50
I-IX	0.38	0.50
I-X	0.46	0.52
II-III	0.30	0.49
II-IV	0.45	0.61
II-V	0.47	0.71
II-VI	0.42	0.55
II-VII	0.45	0.62
II-VIII	0.27	0.49
II-IX	0.42	0.48
II-X	0.60	0.53
III-IV	0.41	0.51
III-V	0.44	0.57
III-VI	0.46	0.61
III-VII	0.18	0.63
III-VIII	0.17	0.46
III-IX	0.26	0.58
III-X	0.16	0.52
IV-V	0.60	0.60
IV-VI	0.49	0.58
IV-VII	0.37	0.62
IV-VIII	0.41	0.44
IV-IX	0.47	0.67
IV-X	0.38	0.46
V-VI	0.33	0.65
V-VII	0.47	0.73
V-VIII	0.38	0.57
V-IX	0.53	0.62
V-X	0.51	0.51
VI-VII	0.28	0.67
VI-VIII	0.33	0.63
VI-IX	0.39	0.58
VI-X	0.43	0.65
VII VIII	0.29	0.56
VII--IX	0.44	0.63
VII-X	0.54	0.49
VIII-IX	0.48	0.40
VIII-X	0.35	0.57
IX-X	0.64	0.48
Average	0.37	0.54

The correlations for the two trials with the 192 children were 0.47 and 0.44 for the two methods of scoring. For the ten trials with the group of 97 children we have a total of 45 coefficients of correlation. They are given in order in Table VIII.

The first column gives the trials correlated and the second and third columns the coefficients for "Consecutive Series" and "Maximum Series" scores respectively. The values are somewhat less than they should be for one reason and somewhat more than they should be for another. As mentioned before, one of the three experimenters changed his method for the third trial so that not one of forty children scored higher than four on the "Consecutive Series." All these correlations, therefore, in which the third trial is considered are lower than we might expect. On the other hand, our series of digits did not go lower than three. A number of children failed in getting this series correct so that their "Consecutive Series" score could only be zero. Since this break occured in the series it did not seem wise to include the zero score in the correlation. The coefficients are, therefore, somewhat higher than we might expect if series of digits down to one had been used.

A collation of the coefficients is shown in Table IX.

TABLE IX.

Coefficients of Correlation	Coefficients of Correlation "Consecutive Series"	"Maximum Series"
0.00–0.09	1	
0.10–0.19	4	
0.20–0.29	4	
0.30–0.39	12	
0.40–0.49	18	10
0.50–0.59	3	18
0.60–0.69	3	15
0.70–0.79		2
Total	45	45

The first column shows them grouped in ranges of ten points. The second and third columns contain the number of the correlations. According to this table the coefficient of correlation for one pair of trials was less than 0.9. For four trials it ranged between 0.10 and 0.19; for another four between 0.20 and 0.29. In only six cases

is the coefficient of correlation for "Consecutive Series" scores greater than 0.50. The average is 0.37. The average coefficient of correlation for the "Maximum Series" score is 0.54. Thirty-three come between 0.50 and 0.69; only two are greater than 0.70. If the memory span were a constant function, one would be justified in expecting a higher correlation than those that are shown. A coefficient less than 0.75 is extremely significant when different functions are being correlated, but correlations of less than that value of the same function on different days leads us to suspect that the apparent validity of the individual test may be seriously questioned.

Conclusion

In our evaluation o fmemory span we have been largely concerned with group or average results. We have determined what the score for children in a great many different problems should be. We have assumed, if a child reaches that average score in one examination, that he has the ability to do it on all occasions. We, therefore, claim that he ranks in intelligence in a certain part of the intelligence scale. On the other hand if in one examination a child fails to make the average score, we have judged him incapable of doing it and ranked him in a lower part of the intelligence scale. Our results indicate that this method of grading needs revision. The performance ability of children is not a static or fixed quantity. It rather varies from month to month, from day to day, or we might even say from hour to hour. A child who is able to make 100% in Spelling one day may make only 75% or 50% in Spelling the same list of words on another day. Children who are able to do 35, 40 or 50 problems a minute one day may do considerably less or considerably more on another day.

When we consider the question from the standpoint of average performance for a group, we do not need to be concerned with this variability. The average score for a group remains approximately constant with the same degree of practice, or increases in fairly constant ratios with similar amounts of practice. Children in one city with the same amount of practice as children in another city give almost identical results. But the individual child may materially change his own position in the rank order. He who is first may not change his position so that he is last but he may change it to such an extent that neither the individual score nor the average

of two scores will give a satisfactory record of his ability or his intelligence, or even of his capability in the specific function tested.

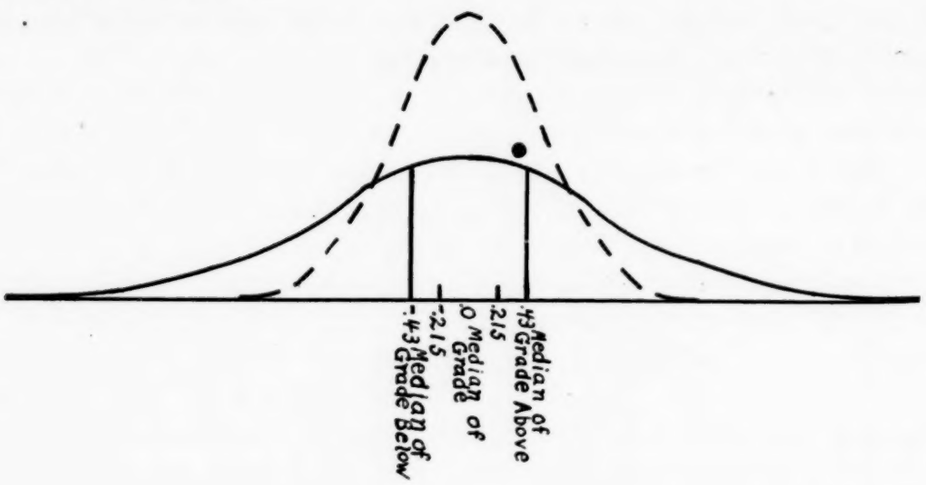
These results may not be considered as conclusive evidence that our mental testing is being done under the handicap of a false assumption. It indicates however, that a revision of our ideas concerning it may be necessary. In the scales of intelligence which have been devised, great weight is placed upon the ability to reproduce numbers and also to reproduce series of words in sentences. The method of giving the test and of scoring probably eliminates some of the variability which we have found, but when the "Maximum Series" score of children can vary as much as six within the range of ten, it does seem that injustice may be done to the child who on a certain day gives his lowest possible score in the memory span. Before the evidence may be considered conclusive, it will be necessary to determine the amount of variability when the method followed is identical with that prescribed in the intelligence scale. There is little doubt, however, that the variability in performance merits careful and serious consideration and demonstrates that the doubt should always be in favor of the child until a wider range of experiment with him has been possible.

COMMUNICATIONS AND DISCUSSIONS

THE MEASUREMENT OF OVERLAPPING*

All who have given standardized tests or have used scales for the measurement of accomplishment of children in different grades have found great variability in each grade and extensive overlapping of grade upon grade. The existence of extreme overlapping in reading ability is reported by Daniel Starch[1]
who writes "it is found that in speed and comprehension 31.8 per cent of the pupils of any grade reach or exceed the median of the next grade above,"and by F. J. Kelly, who reports[2]
that "The reading ability possessed by the median child in any one grade is superior to the ability possessed by at least one third of the children in the class above him and is no better than the reading ability possessed by at least a third in the class below him."

Assuming a normal distribution of scores in reading ability and calling the standard deviation of such scores σ, this statement would imply, as may be determined by reference to a table of the frequencies in a normal distribution, that one grade median is .43 σ distant from the next. The situation is pictured in the accompaning full-line graph.



Let us assume that the score in the reading test is a valid criterion of ability, and determine the per cent of pupils who would be better classified if they were in the grade above or the grade below. Any pupil making a score greater than that which corresponds to .215 σ

*This communication was printed in the April, 1919, issue of the JOURNAL. Owing to an oversight the proof was not read, and the formulae were quite unintelligible. In order to make such amends as are possible, the article is reprinted here. EDITOR.

[1]DANIEL STARCH. *Educational Measurements.* p. 42.

[2]*The Kansas Silent Reading Tests.* J. ED. Psyc. v. 7 No. 2, 1916.

is closer to the median accomplishment of the grade above than to that of the grade in which he is located, and therefore should be transferred to the grade above. The percentage of cases in a normal distribution lying above .215 σ is 41.5, and similarly 41.5 lie below—.215 σ. Accordingly, by this criterion, 17 per cent are properly and 83 per cent improperly classified. As, thruout a number of grades, reading ability is presumably the most important single determinant in classification, so extreme a finding as this may well be questioned and carefully considered before finally accepted.

We may approach this matter from another point of view. If we give a five minute test in solving puzzles we are testing ability to react to novel situations and may call our test a reasoning ability test. From data at hand the writer knows that enormous overlapping would be found with such a test. The reason is here quite obvious. The puzzle ability of an individual may be quite identical with his reasoning ability, but it cannot be determined in five minutes. The probable error of a five minute determination of puzzle ability age is very likely three or four years. The great overlapping in this case is seen to be a consequence of the unreliability of the test, and not the trivial nature of the function tested or of the real overlapping of talent from grade to grade.

To a degree the same situation maintains in the case of every measure of overlapping, unless it be in dealing with such measures as height or weight which can be "exactly" measured, and the unreliability of measures must be known before significant estimates of overlapping can be made.

The writer has found in the case of two classes of about thirty, that the reliability coefficient (the extent to which the test correlates with a second similar test) of both the Starch Reading test and the Kansas Silent Reading test was between .2 and .3. More extensive determinations of these reliability coefficients would probably not change the values appreciably, so it is clear how it has happened that such great overlapping was found.

Knowing the reliability coefficient, it is possible to determine what the overlapping would be if measured by means of a very large number of such tests. We will call this the true overlapping with respect to the ability tested. If the scores of the members of a class are the average scores based upon a very large number of tests, all similar to each other in difficulty and in what each tests, the scores will truly represent ability and whatever overlapping is then found will be overlapping in true ability. Consider the dotted curve shown to represent the distribution of scores based upon averages of a large number of tests. If the grades received in the first test are designated by x_1 (x_2 in the second test, x_3 in the third, x_n in the n'th), where x_1 is a deviation from the group mean, then the full line curve represents the distribution of grades x_1 received

by the members of the class (consider N to be the number of individuals), and the dotted curve represents the distribution of the N average grades, each of which is of this type, $\frac{x_1 + x_2 \cdots + x_n}{n}$. We wish to know the spread or dispersion of these latter measures and will therefore calculate their standard deviation:

$$\sigma^2 \text{ of } \left(\frac{x_1 + x_2 + \cdots x_n}{n}\right) = \frac{\Sigma (x_1 + x_2 + \cdots x_n),^2}{nN}$$ in which Σ indicates a summation of N terms.

$$= \frac{\Sigma x_1^2 + \Sigma x_2^2 + \cdots \Sigma x_n^2 + 2\Sigma x_1 x_2 + 2\Sigma x_1 x_3 + \cdots 2\Sigma x_1 x_n + 2\Sigma x_2 x_3 + \cdots}{n \quad N}$$

Since according to hypothesis the tests which are averaged are all similar they all have the same standard deviation and correlate with each other to the same amount; that is to say

$\Sigma x_1^2 = \Sigma x_2^2 = \cdots = N\sigma_1^2$, according to the usual definition of standard deviation, and

$\Sigma x_1 x_2 = \Sigma x_1 x_3 = \cdots = N r \sigma_1 \sigma^2$, according to the usual definition of a product-moment coefficient of correlation.

In this last equation r is the correlation between one test and a second similar one. In the equation given there are n terms such as Σx_1^2, and $\frac{n(n-1)}{2}$ terms such as $\Sigma 2 x_1 x_2$, so that we have

$$\sigma^2 \text{of} \left(\frac{x_1 + x_2 + \cdots x_n}{n}\right) = \frac{nN\sigma_1^2 + n(n-1)Nr\sigma_1^2}{nN} = \sigma_1^2 \left(\frac{1}{n} + r - \frac{1}{n} r\right)$$

As n becomes indefinitely large the expression in paranthesis becomes r, so that, representing the standard deviation based upon a very large number of tests, *i. e.* the true standard deviation, by σt we have:

$\sigma_t = \sigma_1 \sqrt{r}$ (Formula giving the standard deviation of the average of an indefinitely large number of similar tests, knowing the standard deviation and coefficient of reliability of the single tests)

In the case of the Kansas Silent Reading tests, taking the reliability coefficient = .25 and the standard deviation σ_1, we have

$$\sigma t = \sigma_1 \sqrt{.25}, \text{ or } \sigma t = .5 \ \sigma_1$$

In the original distribution of scores it was necessary to go out from the mean a distance of $.215\sigma$ to reach a score half way between the median scores of two grades. It will accordingly be necessary in terms of the distribution of true scores to go out $\frac{.215}{.5}$ (= .43) standard

deviations to reach the point midway between grades and to go out .86 σ to reach a neighboring grade median. Therefore according to the overlapping reported by F. J. Kelly there is 33 per cent, instead of 17 per cent, of correct classification and, referring to the frequencies given in a normal distribution, 19.5, instead of 33 per cent, exceed or fall short of the median accomplishment of the grade above or below.

All measures of overlapping thus far reported are incorrect as indicative of the numerical amount of overlapping in true ability The correction in the measure of overlapping here derived is very easy to apply; a table of the frequencies in a normal distribution and the coefficient of reliability being the only new data required.

It is seen that great overlapping is as much an indictment of the test as it is the adequacy of the classification, unless it has been determined that the test itself is highly reliable. Three factors make for overlapping: (a) unreliability of tests; (b) unimportance of function tested; and (c) failure to properly classify, and only in so far as the first two are known not to be the causes can improper classification be established.

TRUMAN L. KELLEY.

Teachers College, Columbia University.

THE USE OF A COMPOSITION SCALE

My experience with the Hillegas Scale has been very different from that of Miss Kate Gordon recorded in the November, 1918, issue of this JOURNAL.

I do not find in my notes complete records of my experiments with the Hillegas Scale, but I do have at hand some results with the Nassau Supplement, and as they are in the main similar they are here given.

A mixed class on Tests and Measurements of 16 students in a summer normal school first rated on the percentage basis a composition, typewritten with all its imperfections. The following day 15 of the same pupils rated the same composition by the Nassau Supplement.

The results of the first rating were arranged into 10 groups so as to be more comparable with the groups of the scale. The results were as follows:

Percentage Ratings		*Nassau Supplement*	
%	Frequency	Score	Frequency
0–10		0	
11–20		1	
21–30		2	
31–40		3	1
41–50	4	4	9
51–60	2	5	5
61–70	8	6	
71–80	1	7	
81–90	1	8	
91–100		9	
	16		15

After a further study of the scale these same students were given a set of 32 seventh grade compositions to rate. Each student had two compositions. After recording the results on the blackboard, the students exchanged sets of papers and then proceeded to rate them again on the basis of the scale. Thus each composition was rated by two students. The results were as follows:

Quality	*First time*	*Second time*
0		
1		
2		
3	5	5
4	13	14
5	8	7
6	4	4
7	2	2
8		
9		
	32	32

In combining the two ratings it was found that—

Both marks the same	16
Difference of 1 step	15
Difference of 2 steps	1

ERNEST C. WITHAM

Southington, Connecticut.

THE CHILD'S USE OF NUMBERS

This paper is an attempt to answer two questions: (1) What arithmetical knowledge and power should be acquired by a child by the end of grade VI? And especially, (2) what should be the content of the arithmetic course in grades IV, V, and VI?

The normal American child is the one considered,—one well brought up in the middle class American home, who reaches grade VII, at about the age of twelve. The needs of such a child will set the standard, whatever be the special provisions required for the abnormal child.

School at best provides an unnatural environment for a child. His out-of-school interests are the best expression of his personality. Accordingly I have gone to the children, who knew me, and asked them to tell me what uses of numbers they make or notice outside of school. The children were members of the Henry L. Pierce School in a choice residence section of the city of Boston. I went to twelve classes of grades IV, V, and VI, and to two immature classes of grade VII; in a most informal and friendly way I got them to give me the information I sought.

THE DATA

I have attempted to tabulate the children's answers. Most of the items reported by a lower grade were also reported by higher grades. I have credited each item to the lowest grade that reported it. Here, then, are the ways in which children of their own accord, use numbers.

A. Games (Regarded by children as of great importance).

Reported first by Grade IV

1. Calling or reading numbers
   - (a) Football signals
   - (b) Spots on dominoes
   - (c) Spots on dice
   - (d) Number on sled to indicate size
2. Counting
   - (a) Marbles and tops
   - (b) To give children time to hide
   - (c) Keeping score
   - (d) Number of "swings" while taking turns
   - (e) Strikes in baseball
   - (f) Number of bounces of ball
   - (g) Number of children in a game
   - (h) While going down hill to estimate length of slide
   - (i) Playing with top counting machine

Reported first by Grade VII

   - (j) Number of runners in a race

Reported first by Grade IV

3. Playing Store

   All uses of numbers given below under "Going to Store" are also made while *playing* store

Reported first by Grade V

4. Playing Bank

   Matches were used to represent money

B. Going to Store (All use of money is of relatively small importance to child)

Reported first by Grade IV

1. Knowledge of quantity (such as pounds) of articles to be bought
2. Number of articles to be bought
3. Adding, to find total cost of articles bought
4. Counting change
5. Reading number on paper money to determine denomination

Reported first by Grade V

6. Knowledge of price of article to be bought

C. Other Instances of Reading (or Recognition) of some one number

Reported first by Grade IV

1. Telephone number
2. Automobile number
3. Policeman's number
4. Number on fire engine
5. Number on stove
6. Number on street car
7. Number on police station
8. Number of G. A. R. Post
9. Number on house (both in writing addresses and to find a certain house)
10. Telling time
11. Date, year, and day of month
12. Reading thermometer
13. Tire pressure guage

Reported first by Grade V

14. Number on trainmen (elevated trains)
15. Number on firemen
16. Number on newspaper boy
17. Number of regiment
18. Number of submarine
19. Number ot pew in church
20. Number of stateroom on boat
21. Number on speed-guage of Victrola
22. Numbers on signal-boards of elevators
23. Numbers in a watch
24. Number by which book is known in library
25. Number of library card
26. Number of record for use on Victrola
27. Numbers in shoes
28. Numbers of floors near doorbell push buttons
29. Number on blinds
30. Number on postage stamps
31. Number on keys and lockers at bathhouses
32. Age of person
33. Numbers in music lessons to indicate what fingers are to be used
34. Date on building
35. Date on monument
36. Numbers gotten by measuring objects
37. Number of any one page of a book

38. Number of page in newspaper to which article is continued from page one
39. Some streets have numbers for names
40. Number as part of name of ruler (Haakon II of Norway)
41. Number of minutes allowed in telephone toll call
42. Page of book at which one stops reading

Reported first by Grade VI

43. Train time in timetable
44. Noting numbers on checks while helping father

Reported first by Grade VII

45. Number on fire alarm box
46. Number on motorman (street car)
47. Numbers on scales (weighing)
48. Number of seat in theater
49. Number of size of rubbers
50. Number on delivery wagon of large store
51. Number on ice cream can
52. Dates in history (Home reading, not for school)
53. Sizes of bits (tools used at home)
54. Total number of pages in book
55. Time when sun rises and sets

D. Other Instances of Counting

Reported first by Grade IV

1. Money belonging to child
2. Trading stamps
3. Postage stamps in child's collection
4. "Checks" good for ice cream at store
5. Books at home belonging to child
6. Stars on flag
7. Stones while digging garden
8. Stars at night

Reported first by Grade V

9. Steps in dancing lessons
10. Days until some holiday
11. Stories in a book
12. Houses on a street
13. Serving of food to supply family at table
14. Time in music lessons and dancing lessons
15. Number of pages read in book
16. Numbering of pages when using typewriter

Reported first by Grade VI

17. Number of Christmas presents
18. Number of guests at table
19. Stripes (as well as stars) on flag

Reported first by Grade VII

20. Money, when selling
21. Stitches while sewing
22. Number of places while setting table

23. Number of times in dumbbell exercises
24. Number of tracks in large railroad station
25. Number of dishes washed
26. Windows, to know how many shades to buy
27. Stripes on policeman's sleeve
28. Newspapers, by boy who is selling them
29 Advertising circulars sent out by father (counted by child)
30. Pieces of carrot for rabbit
31. Rabbits, when shut in
32. Strokes of ax while cutting wood
33. Tulips, and other flowers
34. Number of articles ironed
35. Number of keys on piano (counted for fun, not a part of lesson)
36. Number of strokes one can swim
37. Number of steps required to cross street
38. Number of letters per minute in signalling (boy scout)
39. Number of strokes while beating carpet

E. Miscellaneous

Reported first by Grade IV

1. Adding items in personal account book

Reported first by Grade V

2. Adding up Larkin Soap order
3. Making change while working in store

Reported first by Grade VI

4. Making out bills at store where boy works

Reported first by Grade VII

5. Timing what is cooking

THE RESULTS

Thus do the children reveal the uses of numbers that everyday life causes them to make. Compare this mass of data with the table of contents of any textbook in arithmetic and notice how limited the children's uses are. *Reading of numbers and counting include nearly every item recorded above.* Again let us remind ourselves that these are children from good homes, sheltered as we schoolmasters shelter our own children. In this study I am seeking to learn what is true of normal children.

Below grade VII, then, there are no needs felt by the child which require the teaching of arithmetic in the school. All he cares for is to count and to read numbers.

What, then, shall we teach?

In the first place we may at once discard *applied* arithmetic in the grades below the Junior High School. Our textbooks may drop the "Practical Problems," because they are as foreign as Chinese to a child's way of thinking.

But do not think that a child dislikes numbers. I find in these younger children a keen delight in the manipulation of pure numbers (abstract numbers). Mechanical drill is now a pleasure.

Grades IV, V, and VI, then, afford the ideal opportunity for drill in pure numbers (that is, abstract numbers, as opposed to applied arithmetic). This number drill should give the child knowledge and power to work the four processes (addition, substraction, multiplication and division) with integers, common fractions (what little is needed) and decimals. The work should be systematic, but not many reasons should be given for the various processes taught; for example, for inverting the divisor in division of fractions. Logical reasons in arithmetic can be easily taught in the Senior High School.

The work with integers and decimals, by the last part of the sixth grade, should be as difficult as a person will meet until he uses an adding machine and a table of logarithms. The Courtis Tests should come to a maximum in grade VI.

SUMMARY

The children in what we now call the elementary school grades make little use of arithmetic in their life outside of school. They like to count and to read numbers, and that is about all. But they have keen delight in the manipulation of pure numbers. Here, then, we have an opportunity to perfect them in integers and in common and decimal fractions. Then comes the Junior High School when the child awakens to the meaning of life. The arithmetic drill in grades IV, V, and VI, establishes the foundation for the mathematics of the Junior High School.

PHILO G. NOON.

Boston, Massachusetts.

THE ATTENTION VALUE OF LECTURING WITHOUT NOTES

It is a matter of common observation that a lecturer more easily holds his audience when speaking without notes than when reading from them, but so far as is known to the writer no attempt has yet been made to submit this accepted fact to a quantitative statement, or to analyze the factors that may be significantly compared in the two methods of delivery. While the relative advantages of speaking and reading statements to an audience obviously vary widely with individual lecturers, there are unquestionably certain general tendencies which, if successfully analyzed, would be found to obtain with a high degree of constancy for a considerable majority of speakers.

Two types of factors must be clearly distinguished. For one thing the lecturer without notes unconsciously varies his spoken words according to the varying responses from his hearers. He continually modifies set statements into conversational form, changes

his method of emphasis to meet the fluctuations of interest that he sees about him, elaborates where he finds that he has not been clear, and forces himself to greater animation when he meets evidence that he has been dull. These subjective changes in individual speakers are probably too diverse to admit of constant formulation, and would not lend themselves easily to experimental measurement.

But aside from the lecturer's sympathetic variations of statement and gesture there is a sheer difference in the audience according to whether it is listening to reading or to speech. This difference is immediately apparent in the uneasy restlessness of any audience the moment that a speaker who has been expected to talk to them suddenly produces a manuscript. This fact has often impressed the writer as he noted his own responses as a member of an audience, and the following class experiments were made in the effort to arrive at a quantitative statement of the difference in question.

The general method was to give identical material to two very similar groups of students, reading from lecture notes to one class, and speaking from memory to the other. Immediately after hearing the material the students were instructed to write out in the completest detail all they could remember of what they had just heard. Then the papers were scored according to the number of ideas correctly remembered.

The two groups of students numbered 39 and 61 respectively. Both were made up of college undergraduates at the University of Minnesota, and all were beginning students in psychology. Fifty-four per cent of each group were women, in actual numbers 21 of the 39 and 33 of the 61. That the general intelligence of the two groups was approximately the same may be indicated by the distribution of the marks in the course in psychology. The group of 39 numbered 5 A, 7 B, 14 C, 9 D, and 4 E grade students. The group of 61 numbered 6 A, 13 B, 25 C, 11 D, and 6 E students.

Two class experiments were performed, one preliminary and one final. For the preliminary experiment the familiar Marble Statue test was used as material for scoring logical memory. The story was spoken from memory to the larger group and read to the smaller group. The distinctive fact in this preliminary trial was that both groups were aware that the material concerned a special test, and attended with an interest in competition for score. The score for the two groups shows no significant difference. Those who heard the material read averaged 40, with a mean variation of 6.2; those who heard it spoken averaged 41.6, with a mean variation of 3.3. Evidently when the audience is attending competitively the matter of reading or speaking becomes of small importance.

The distinctive feature of the final experiment was that the material for use in the memory test was introduced as part of the regular lecture in psychology, and the students were in no way led to suppose that any peculiar importance attached to it. The material used was a 184 element statement concerning the life and scientific work of Helmholtz, and came at a time when both classes

had just finished the study of visual and auditory sensations in Pillsbury's "Fundamentals of Psychology." The unusual importance of Helmholtz to the subject of psychology was made the justification for introducing a certain amount of biographical material while summarizing the work in audition and vision.

Exactly five minutes were given to the reading and to the speaking, and a long previous practice had assured practically the same kind of emphasis, pauses, and changes of intonation, as heard by the two classes. A person standing outside the room would have had great difficulty in distinguishing which statement was read and which one spoken. But the scores in this test were noticeably different for the two groups. Reading gave an average score of 49.6, with a mean variation of 14.4, and speaking an average of 67.5, with a mean variation of 15.7.

This average difference of 36 percent as contrasted with the negligible difference in the preliminary experiment seems to indicate that the disposition of an audience to give attention is one third greater for spoken utterance than for reading. When any ulterior motive, such as competition, comes into play, attention is about equally effective for reading and speaking, at least for a short time.

It may be concluded that a speaker who reads his address has to assume an unusual earnestness of purpose in his hearers, a much greater earnestness than normally exists in the classroom. And when this sheer difference of 36 percent in the original disposition of the audience is taken in conjunction with the still larger differences due to the reciprocal exchange between speaker and audience, it is hard to escape the conviction that the lecture method, in the sense of the reading of notes to hardly persuaded students, is one of the most dubious features of present day method in college teaching. To the extent that a teacher makes use of notes in the classroom, to that extent he is incurring the danger of reducing his real function in the college to the mere marking of class attendance.

HENRY T. MOORE.

Damouth College.

The Journal of Educational Psychology

INCLUDING EXPERIMENTAL PEDAGOGY, CHILD PHYSIOLOGY AND HYGIENE, AND EDUCATIONAL STATISTICS

Published monthly except June to August by
WARWICK & YORK, INC.
HERSHEY, PA. BALTIMORE, MD.

Board of Editors
WILLIAM CHANDLER BAGLEY, CARL E SEASHORE,
J. CARLETON BELL, Managing Editor, GUY MONTROSE WHIPPLE.

COLLABORATORS

Sydney Alrutz,	H. H. Goddard,	Will S. Monroe,	L. William Stern
J. R. Angell,	Joseph Jastrow	E. C. Moore,	G. M. Stratton,
J. W. Baird,	G. E. Johnson,	M. V. O'Shea,	G. D. Strayer,
Bird T. Baldwin,	E. A. Kirkpatrick,	W. B. Pillsbury,	E. L. Thorndike,
W. Van Dyke Bingham,	E. H. Lindley,	S. H. Rowe,	E. B. Tichener,
E. F. Buchner,	Anna J. McKeag,	Sante de Sanctis,	J. B. Watson,
W. H. Burnham,	D. P. MacMillan,	E. G. Sanford,	W. H. Winch,
Ed. Claparede,	Adolf Meyer,	W. S. Small,	H. K. Wolfe,
S. S. Colvin,	J. B. Miner,	E. D. Starbuck,	A. H. Yoder.
W. F. Dearborn,			

Manuscripts for publication, books or other material for review, and news items of interest to educational psychologists should be addressed to The Managing Editor, J. Carleton Bell, 1032A Sterling Place, Brooklyn, N. Y.

NOTICE TO SUBSCRIBERS. Notify the publishers, Warwick & York, Inc., of any change of address; as copies lost in transit because of subscriber's removal to another address will not be duplicated

THE JOURNAL is issuing a series of Monograph Supplements for the publicaiion of more extensive researches in educational psychology. For details prospective contributors may apply to H. E. Buchholz, 10 East Center St., Baltimore, Md.

The price of the JOURNAL is $3.00 per year of nine monthly issues, payable in advance. Single issues are 50 cents each. Back volumes are $5.00 in original wrappers and $6.25 bound in library buckram. Foreign postage is 20 cents extra for Canada and 30 cents for other countries in the Postal Union.

Subscriptions and correspondence concerning business matters should be sent to Warwick & York, Inc.

VOLUME X NOVEMBER NUMBER 9

EDITORIAL

THE NATIONAL RESEARCH COUNCIL

The recent World War was preeminently a scientific war. It was a clash of brains as much as a contest of brawn. It was due to the scientific organization and conservation of their resources that the Germans were able to come so near to their goal, and were able to hold out so long after the hope of a smashing success had faded. Trench warfare, the airplane, the long distance gun, the submarine, the tank, are merely the outstanding examples of thousands of scientific devices to sustain the cause and overthrow the enemy. It is distinctly to the credit of the leaders of this country that the significance of scientific organization was realized at the very outset and that every effort was made to mobilize speedily and effectively the scientific resources of the country. This was brought about through the National Academy of Sciences, which at the request of the President established the National Research Council, an organization representing the chief scientific activities of the country. An Executive Order of May 11, 1918, indicates the functions and duties of the Council, in part, as follows:

"In general, to stimulate research in the mathematical, physical and biological sciences, and in the application of these sciences to engineering, agriculture, medicine, and other useful arts, with the object of increasing knowledge, of strengthening the national defense, and of contributing in other ways to the public welfare.

"To survey the larger possibilities of science, to formulate comprehensive projects of research, and to develop effective means of utilizing the scientific and technical resources of the country for dealing with these projects.

"To promote cooperation in research, at home and abroad, in order to secure concentration of effort, minimize duplication, and stimulate progress."

Educational psychology was actively represented on the Council from its very beginning. Already a group of psychologists had rendered signal service in devising tests of intelligence to be applied to the National Army. Early in 1919 a special committee of the National Research Council consisting of Dr. R. M. Yerkes, chairman, Dr. M. E. Haggerty, of the University of Minnesota, Dr. L. M. Terman, of Stanford University, Dr. E. L. Thorndike, of Teachers College, Columbia University, and Dr. G. M. Whipple, of the University of Michigan, secured from the General Education Board a grant of $25,000 for the purpose of working out a series of tests of intelligence for use with school children, similar to those already employed in the army. After detailed study the committee selected some twenty tests and gave these a careful trial on five thousand children. From these tests two series were selected which seemed to give the most satisfaction, and in order to perfect the procedure and to still further check up on the results the tests were tried on several thousand more children. As a result the committee has now before it and will shortly be able to offer to the country two series of group tests of general intelligence of a high degree of reliability that have been carefully tested out on large groups of children, and that give as satisfactory a measure of intelligence as any scale now available. The practical value of such a group scale is tremendous. Schoolmen generally have come to see the advantages of the Binet measurements of intelligence, but the necessity of examining each pupil individually has put such a time cost on the use of the scale as to restrict its employment to special cases. With such a group scale fifty or a hundred children may be examined in an hour, and the only further labor is found in scoring and tabulating the results. We may confidently expect, therefore, that the National Research Council tests will make intelligence testing a routine measure in the schools, and this will be the entering wedge for other studies of pupils' abilities and attainments, and for a radical adjustment of school work to meet the needs of pupils.

J. C. BELL.

NOTES AND NEWS

The New York Society for the Experimental Study of Education held its first meeting of the year in Washington Irving High School, Friday evening, October 10. The general topic of the Meeting was "Problems in the Teaching of English." The following papers were read: "Efficiency of Student Correction of Composition Papers," J. C. Tressler, Newton High School; "The Relative Usefulness of Composition Scales," Sterling A. Leonard, Lincoln School; "Studies in the Appreciation of English Poetry," Professor M. R. Trabue, Teachers College.

The Society for the Promotion of Engineering Education has appointed a committee on intelligence tests for admission to engineering schools. This committee has prepared a plan for cooperative research on admission, and contemplates a series of reports on the predictive value of various tests with respect to subsequent scholarship. A battery of six tests (arithmetic, algebra, geometry, intelligence, physics, and technical information) was prepared for use this fall, and the results of these will be worked over and reported on next summer.

The University of Illinois Bureau of Educational Research has issued an enlarged list of educational tests that it is prepared to furnish, and promises a descriptive list giving detailed information and a brief bibliography of each test. The latter list will be sold at twenty cents per copy.

The National Committee on Mathematical Requirements has received a grant of $16,000 from the General Education Board for the enlargement of the scope of its work. It proposes to make a careful and thorough study of the teaching of mathematics, and to report on various reforms that have been advocated. Professor J. W. Young, of Dartmouth College, and Mr. J. A. Foberg, of the Crane Technical High School, Chicago, have secured leaves of absence from their institutions and are devoting their entire time to the work.

Tests of intelligence are being increasingly used to determine fitness for school work. The school authorities of Lawrence, Kansas, have decided to admit to the schools those children only five and one-half years old who show a mental age of six years, and at the University of Chicago kindergarten children who stand high in intelligence are found to do well in the work of the first grade. These are indications of the growing disposition to study the child and to fit the school activities to his needs and capacities.

In an address before the London Modern Language Association Mr. E. Allison Peers advocated the formation of a department of educational experiment and research whose function it should be to carry on cooperative investigations and publish data on all kinds of innovations in method. After the address a large number of teachers expressed their desire to take part in the experiment.

The Alabama legislature has passed the Alabama Mental Deficiency bill appropriating $200,000 for the establishment of the Alabama Home for Mental Inferiors at Tuscaloosa.

Warwick and York announces the publication of an Introductory Psychology for Teachers by Professor Edward K. Strong, of the Carnegie Institute of Technology. This text has been developed along novel lines, and in the opinion of many who have examined the work will better meet the needs of prospective teachers than any book now on the market.

At Harvard University William McDougall, of Oxford University, has been elected professor of psychology to fill the vacancy caused by the death of Hugo Munsterberg. Professor McDougall will take up his work at the beginning of the next academic year. Dr. Herbert S. Langfeld has been appointed director of the psychological laboratory, and Drs. L. T. Troland and Floyd H. Allport have been appointed instructors in psychology.

Professor Daniel Starch, of the University of Wisconsin, is on leave of absence for the first half of the present year. He is giving a three-hour course of lectures during the semester at Harvard University.

Mr. Leon O. Smith, director of research in the Omaha Public Schools, has been advanced to the position of assistant superintendent.

Dr. Walter S. Monroe, of Indiana University, has been made assistant director of the Bureau of Educational Research of the University of Illinois.

J. M. O'Gorman, lecturer in education at the University of Illinois, has accepted the position of professor of education and psychology in the Montana State College.

PUBLICATIONS RECEIVED

ERNEST R. BRESLICH. *Correlated Mathematics for Junior Colleges.* Chicago: The University of Chicago Press, 1919. Pp. xiii, 301. $1.25.

This constitutes the author's fourth-year text in correlated mathematics. Instead of the traditional algebra in the first year, plane geometry in the second, solid geometry and trigonometry in the third, and analytical geometry in the fourth, this series of texts presents the simpler aspects of the whole subject of mathematics in the first year, and proceeds with increasing complexity through the subsequent three years. The series is constructed throughout from the standpoint of the psychology of the pupil, and gives a well-rounded conception of elementary mathematics.

B. R. BUCKINGHAM. *Extension of the Ayres Spelling Scale.* Bureau of Educational Research, University of Illinois, Urbana, Illinois. 1919. Ten cents.

This extension of the Ayres Scale will distinctly increase its value and usefulness. The 1000 original Ayres words are retained and printed in Roman type. To these 505 new words have been added in italics. The new words were selected from spellers, and their difficulty determined by tests in several states. Most of them fall at the upper end of the Ayres Scale, and it was found desirable to extend the scale six columns to the right. In addition scores for the ninth grade are included for all columns to the right of P. This is the kind of work that will be of real service to educational procedure. The Scale should be still further extended to include perhaps five thousand words.

STUART A. COURTIS. *The Gary Public Schools. Measurement of Classroom Products.* New York: General Education Board, 1919. Pp. xxii, 532. Thirty cents.

This is the final volume of the report on the survey of the Gary schools. It is not only a detailed account of what is unquestionably the most accurate experimental survey of a school system ever made, but is also a critical and constructive discussion of the whole subject of educational measurements. The specific studies reported are handwriting (scored by the Ayres scale), spelling (compared with the Ayres standards), arithmetic (measured by the Courtis tests, Series B, and the Cleveland Survey tests), composition (evaluated by the Hillegas scale), and reading (measured by the Gray Oral Reading scale and the Kansas Silent Reading tests). No student of educational measurements can afford to be without this volume.

KARY C. DAVIS. *Horticulture.* Philadelphia: J. B. Lippincott Company, 1919. Pp. vii, 416. $1.75.

Here is a splendid text for high and normal schools, which combines instruction in gardening, orcharding and small fruit culture, covering in a comprehensive yet wholly practical manner the entire subject of horticulture. The abundant illustrations are carefully planned to supplement and clarity the discussions of the text, and the questions at the end of each chapter stimulate the student to reflective thinking. The book on account of its simple and easy style will also make a strong appeal to the practical home gardener.

ELBERT KIRTLEY FRETWELL. *A Study in Educational Prognosis.* New York: Teachers College, Columbia University Contributions to Education, No. 99, 1919. Pp. 55.

The chief problem in this study was what tests are of most value for educational prognosis? The work was carried on under the supervision of Professor Thomas H. Briggs in the Speyer School, New York City. Eleven tests were given to seventy-four 6 B boys at the beginning and again at the end of the school year, and the results of these were compared with the class records and the judgments of teachers as to school abilities. The author concludes that the tests afford a better means of educational prognosis than do previous school marks or teachers' estimates. The order of value of the tests is reading, visual vocabulary, opposites, spelling, completion, arithmetic, easy directions, mixed relations and composition.

WALTER S. HUNTER. *General Psychology.* Chicago: University of Chicago Press, 1919. Pp. xiii, 351. $2.00.

By the term General Psychology the author intends to indicate his conviction that the field of psychology is much broader than normal adult psychology, and that an introductory work should serve to give the student a bird's eye view of the entire subject. One-third of the book is, therefore, devoted to the consideration of the fields of psychology, including animal psychology, individual and applied psychology, abnormal psychology, and social and racial psychology. In the remainder of the book, the section on normal adult psychology, the effort is made to bring out the essential results of experimental psychology by means of a somewhat detailed account of representative experimental studies. There are many excellent figures of the nervous system and the sense organs, and several graphs showing the results of experiments. As a concise and well balanced introductory text the book deserves careful consideration.

BUFORD JOHNSON. *Practice Effects in a Target Test—A Comparative Study of Groups Varying in Intelligence.* Reprinted from the Psychological Review, Vol. 26, 1919. 300-326.

At the New York State Reformatory for Women three groups of inmates were formed on the basis of intelligence as shown by the Stanford-Binet tests and the Yerkes-Bridges scale. The average I. Q. for each group was 99, 73, and 56. Twenty-seven daily practice periods of ten trials with each hand were given to each group. The high intelligence group had the highest initial and final average, the medium group was second, and the low group last. The medium group, however, showed more improvement than either the high or the low group. The individual learning curves suggest temperamental differences as more significant than differences in intelligence per se.

WILFRID LAY. *The Child's Unconscious Mind.* New York: Dodd, Mead and Company, 1919. Pp. 329. $2.00.

In this book the author tries to show the relations of psychoanalysis to education. The entire argument is based on the assumed activities of the unconscious (subconscious and co-conscious are used as synonyms) mind. This is nowhere clearly explained but under the name of unconscious wish, unconscious inference, unconscious memory, etc., is made responsible for most of our attitudes and strivings. Thus the author goes on about the interplay between the conscious and the unconscious,

identifications, projections, introjections, compensations, sublimations, and the whole stock in trade of the psychoanalysts, drawing his illustrations from child behavior and applying his conclusions to educational practice. All this, shot through as it is with the usual crimson thread of sex reference, impresses the critical psychologist as an erotic phantasmagoria, punctuated here and there with shrewd observations on child life and pertinent suggestions for school room procedure.

Nevada, Report of a Survey of the University of. Washington: Bureau of Education, 1917, No. 19. Pp. 184. Twenty-five cents.

The report refers to the suspicion and distrust that are rife among the people of Nevada regarding the policies of the University, examines its present organization and equipment, commends much that is found, and makes suggestions for betterment. Among the recommendations the increase of attention to the training of teachers for Nevada schools is given a prominent place.

HELEN RICH NORTON. *Department Store Education.* Washington: Bureau of Education, 1917, No. 9. Pp. 79. $.15.

An account of the course of study, methods of teaching and follow-up work developed at the Boston School of Salesmanship under the direction of Lucinda Wyman Prince. There is also a chapter on the teaching of salesmanship in the public schools.

STEPHEN PAGET. *Adolescence.* New York: E. P. Dutton and Company, 1919. Pp. 46.

The address of an eminent physician to Oxford University Extension Students in 1917. A splendid instance of the way educated English people are apt to talk around this subject without ever coming to grips with it.

Personnel Work in the United States Army. Washington: Adjutant General's Office, 1919. Pp. 15.

This pamphlet describes the classification of soldiers according to their industrial qualifications, lists the trade tests developed for determining industrial skill, and presents the rating scale for the classification and rating of officers.

FRANCIS G. PEABODY. *Education for Life.* New York: Doubleday, Page and Company, 1918. Pp. xxiv, 393. $2.50.

The story of Hampton Institute, told in connection with the fiftieth anniversary of the school. It tells of the negro in the Civil War, in the reconstruction period, and in the period of industrial expansion in the South. It recounts the heroic efforts of General Armstrong to provide educational opportunities for the negro, and to raise money for the support of Hampton Institute. The book is profusely illustrated and is provided with valuable statistical appendices. Written as it was by the vice-president of the board of trustees, it may be considered the official history of the first institution for the higher education of the negroes.

JOSEPH PETERSON AND QUENTIN J. DAVID. *The Psychology of Handling Men in the Army.* Minneapolis: Perine Book Company, 1918. Pp. v, 146.

This book is an attempt by a professor of psychology and a lieutenant in the army to apply the psychology of group activities to the problems of handling soldiers. It consists largely of anecdotes drawn from military and athletic life illustrating the values of competition, play, teamplay, leadership, habit and discipline, loyalty, and the principles of learning in dealing with recruits.